

Modelli Markoviani nell'Acquisizione del Linguaggio

Processi Stocastici e Apprendimento delle Sequenze Fonologiche e Morfologiche nei Bambini

Fondamenti Teorici e il Problema dell'Acquisizione Linguistica

L'acquisizione del linguaggio da parte dei bambini rappresenta uno dei fenomeni cognitivi più straordinari e misteriosi. In un periodo relativamente breve, tipicamente dai dodici ai trentasei mesi di vita, i bambini passano dall'emettere semplici balbettii a produrre frasi grammaticalmente complesse nella loro lingua nativa, senza istruzione formale esplicita e spesso in condizioni di input linguistico imperfetto o frammentario. Questa capacità di estrarre pattern regolari da dati rumorosi e incompleti, di generalizzare da esempi limitati a una competenza produttiva illimitata, e di convergere rapidamente su sistemi grammaticali condivisi con la comunità linguistica circostante, ha affascinato linguisti, psicologi e scienziati cognitivi per generazioni.

Il problema centrale dell'acquisizione del linguaggio, articolato influentemente da Noam Chomsky negli anni Cinquanta e Sessanta, è quello che viene chiamato il "problema logico dell'acquisizione" o "poverty of the stimulus argument". L'argomento può essere formulato così: i bambini sono esposti a un insieme finito di enunciati, molti dei quali contengono errori di performance, false partenze, e frammentazioni. Eppure, da questo input impoverito, i bambini convergono su grammatiche che consentono loro di comprendere e produrre un insieme potenzialmente infinito di frasi nuove. Come è possibile questo salto dall'esperienza limitata alla competenza generalizzata?

La risposta tradizionale nella linguistica generativa è stata postulare l'esistenza di una Grammatica Universale innata, un insieme di principi e parametri biologicamente determinati che vincolano lo spazio delle grammatiche possibili, rendendo il problema dell'apprendimento computazionalmente trattabile. Secondo questa visione, i bambini non devono indurre la grammatica della loro lingua da zero, ma piuttosto devono semplicemente "impostare i parametri" di una grammatica già sostanzialmente presente nel loro patrimonio genetico.

Tuttavia, a partire dagli anni Novanta, una prospettiva alternativa ha guadagnato forza, basata su modelli computazionali e statistici dell'apprendimento linguistico. Questa prospettiva, radicata nella teoria della probabilità e nei processi stocastici, suggerisce che meccanismi di apprendimento statistico dominio-general, combinati con vincoli cognitivi e percettivi, potrebbero essere sufficienti per spiegare molti aspetti dell'acquisizione linguistica senza ricorrere a conoscenze grammaticali innate specifiche.

I modelli markoviani, e più in generale i processi stocastici, giocano un ruolo centrale in questa prospettiva alternativa. Questi modelli trattano il linguaggio come una sequenza di elementi (fonemi, sillabe, morfemi, parole) dove la probabilità di osservare un dato elemento in una posizione dipende dagli elementi che lo hanno preceduto. Attraverso l'esposizione ripetuta a sequenze linguistiche nell'input ambientale, il bambino può apprendere queste probabilità di transizione e usarle per segmentare il flusso continuo del parlato, identificare unità linguistiche discrete, e indurre regolarità fonologiche e morfologiche.

L'applicazione di modelli markoviani all'acquisizione del linguaggio non è meramente un esercizio di modellazione computazionale, ma riflette ipotesi empiriche testabili sulla natura dei meccanismi di apprendimento infantile. Studi sperimentali hanno dimostrato che neonati e bambini piccoli sono straordinariamente sensibili a pattern statistici nell'input linguistico. Possono tracciare probabilità di transizione tra sillabe dopo solo pochi minuti di esposizione a lingue artificiali. Questa sensibilità alle statistiche sequenziali emerge molto precocemente, addirittura prima della nascita, ed è stata documentata non solo per stimoli linguistici ma anche per stimoli visivi e uditivi non linguistici, suggerendo un meccanismo di apprendimento statistico dominio-generale.

La prospettiva markoviana fornisce un framework formale preciso per modellare e testare ipotesi su come questa sensibilità statistica contribuisce all'acquisizione di componenti specifici del linguaggio. In particolare, i modelli markoviani si sono rivelati particolarmente potenti per comprendere:

Segmentazione del parlato: Come i bambini identificano confini di parola in un flusso continuo di suoni senza pause esplicite tra parole.

Acquisizione fonologica: Come i bambini apprendono l'inventario dei suoni distintivi (fonemi) della loro lingua e le regole che governano le loro combinazioni (fonotassi).

Morfologia: Come i bambini scoprono le unità minimali di significato (morfemi) e apprendono le regolarità morfologiche come formazione del plurale, coniugazione verbale, e derivazione lessicale.

Transizioni sintattiche: Come i bambini iniziano a tracciare pattern di adiacenza tra categorie di parole, precursori della conoscenza sintattica più astratta.

Prima di esplorare in dettaglio come i modelli markoviani illuminano questi aspetti dell'acquisizione, è necessario stabilire un framework matematico rigoroso per comprendere cosa sono i processi markoviani e come possono essere applicati al linguaggio.

Processi Stocastici e Sequenze Temporali

Un processo stocastico è una collezione di variabili casuali indicizzate dal tempo o da qualche altra dimensione ordinata. Formalmente, un processo stocastico discreto è una sequenza $\{X_1, X_2, X_3, \dots\}$ dove ogni X_i è una variabile casuale che assume valori in qualche spazio di stati S . Nel contesto del linguaggio, lo spazio degli stati potrebbe essere l'insieme dei fonemi, l'insieme delle sillabe, l'insieme delle parole, o qualche altra unità linguistica.

Per esempio, consideriamo una frase come una sequenza di parole. Se rappresentiamo ogni parola come uno stato, allora una frase è una realizzazione particolare del processo stocastico: $X_1 = \text{"il"}$, $X_2 = \text{"gatto"}$, $X_3 = \text{"dorme"}$, e così via. Il processo stocastico codifica non solo quali sequenze sono osservate, ma anche quali sequenze sono probabili e quali sono improbabili.

La distribuzione di probabilità completa di un processo stocastico specifica, per ogni possibile sequenza finita, la probabilità di osservare quella sequenza:

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$$

Per processi generali, questa distribuzione congiunta può essere estremamente complessa. La forza dei modelli markoviani risiede nella loro assunzione semplificatrice che riduce drasticamente la complessità pur mantenendo potere espressivo significativo.

Catene di Markov: Definizione Formale

Una catena di Markov è un processo stocastico che soddisfa la proprietà di Markov: il futuro è indipendente dal passato dato il presente. Formalmente, per ogni n e per tutti gli stati $x_1, x_2, \dots, x_{n+1}$:

$$P(X_{n+1} = x_{n+1} \mid X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = P(X_{n+1} = x_{n+1} \mid X_n = x_n)$$

In parole: la probabilità di transizione al prossimo stato dipende solo dallo stato corrente, non dalla storia completa di come si è arrivati allo stato corrente. Questa è chiamata proprietà markoviana di primo ordine.

Più generalmente, una catena di Markov di ordine k soddisfa:

$$P(X_{n+1} = x_{n+1} \mid X_1 = x_1, \dots, X_n = x_n) = P(X_{n+1} = x_{n+1} \mid X_{n-k+1} = x_{n-k+1}, \dots, X_n = x_n)$$

cioè, il prossimo stato dipende solo dai k stati precedenti. Nel contesto linguistico, questo corrisponde a modelli n -gram dove $n = k + 1$.

Una catena di Markov è completamente specificata da:

1. **Spazio degli stati** S : L'insieme di tutti i possibili stati che il processo può assumere.
2. **Matrice di transizione** P : Una matrice $|S| \times |S|$ dove l'elemento $P[i,j]$ specifica la probabilità di transizione dallo stato i allo stato j :

$$P[i,j] = P(X_{n+1} = j \mid X_n = i)$$

La matrice di transizione è stocastica, il che significa che ogni riga somma a 1:

$$\sum_j P[i,j] = 1 \text{ per ogni } i$$

3. **Distribuzione iniziale** π : Un vettore di probabilità che specifica la distribuzione dello stato iniziale:

$$\pi[i] = P(X_1 = i)$$

$$\text{con } \sum_i \pi[i] = 1$$

Per catene di Markov tempo-omogenee, le probabilità di transizione non cambiano nel tempo: $P(X_{n+1} = j \mid X_n = i)$ è la stessa per ogni n . Questa è un'assunzione semplificatrice comune, sebbene si possano considerare anche catene non-omogenee dove le probabilità di transizione evolvono.

Proprietà Fondamentali delle Catene di Markov

Le catene di Markov possiedono diverse proprietà matematiche importanti che sono rilevanti per la loro applicazione all'acquisizione linguistica.

Probabilità di sequenze: Data una catena di Markov, la probabilità di osservare una particolare sequenza $x_1, x_2, \dots, x_n$ si fattorizza come:

$$\begin{aligned} P(X_1 = x_1, \dots, X_n = x_n) &= \pi[x_1] \times P[x_1, x_2] \times P[x_2, x_3] \times \dots \times P[x_{n-1}, x_n] \\ &= \pi[x_1] \times \prod_{i=1}^{n-1} P[x_i, x_{i+1}] \end{aligned}$$

Questa fattorizzazione semplice è cruciale per calcoli efficienti.

Equazioni di Chapman-Kolmogorov: La probabilità di transizione in n passi, denotata $P^{(n)}[i, j] = P(X_{n+1} = j \mid X_n = i)$, soddisfa:

$$P^{(n)} = P^n$$

dove P^n indica la matrice P moltiplicata per se stessa n volte. Questo permette di calcolare probabilità di transizione a lungo termine.

Classificazione degli stati: Gli stati di una catena di Markov possono essere classificati come:

- **Ricorrenti:** Stati a cui il processo ritorna con probabilità 1
- **Transienti:** Stati che potrebbero essere visitati ma non necessariamente rivisitati
- **Assorbenti:** Stati da cui non si può uscire ($P[i, i] = 1$)
- **Comunicanti:** Stati i e j sono comunicanti se esiste un cammino di probabilità positiva da i a j e viceversa

Una catena di Markov è **irriducibile** se tutti gli stati sono comunicanti tra loro. È **aperiodica** se non esistono cicli di lunghezza fissa che vincolano quando uno stato può essere rivisitato.

Distribuzioni stazionarie: Una distribuzione di probabilità π^* sugli stati è stazionaria se:

$$\pi^* = \pi^* P$$

cioè, se la catena si trova nella distribuzione π^* al tempo n , rimarrà in quella distribuzione al tempo $n+1$. Per catene finite, irriducibili e aperiodiche, esiste una unica distribuzione stazionaria π^* che è anche la distribuzione limite:

$$\lim_{n \rightarrow \infty} P^{(n)}[i,j] = \pi^*[j]$$

indipendentemente dallo stato iniziale i . Questa proprietà di convergenza è importante per comprendere il comportamento a lungo termine del processo.

Entropia e tasso di entropia: L'entropia di una distribuzione sugli stati misura la sua incertezza:

$$H(\pi) = -\sum_i \pi[i] \log \pi[i]$$

Per una catena di Markov stazionaria, il tasso di entropia (entropy rate) è:

$$H_{\text{rate}} = -\sum_i \pi^*[i] \sum_j P[i,j] \log P[i,j]$$

e rappresenta l'incertezza media per simbolo in sequenze lunghe generate dal processo. Il tasso di entropia è una misura della complessità o imprevedibilità intrinseca del processo.

Modelli N-Gram e Linguaggio

Nell'elaborazione del linguaggio naturale, i modelli n-gram sono un'applicazione diretta delle catene di Markov. Un modello n-gram predice la probabilità di una parola basandosi sulle $n-1$ parole precedenti.

Un **modello unigram** (1-gram) assume che ogni parola sia indipendente:

$$P(w_1, w_2, \dots, w_n) = \prod_{i=1}^n P(w_i)$$

Questo è il modello più semplice ma anche il meno realistico, poiché ignora completamente il contesto.

Un **modello bigram** (2-gram) corrisponde a una catena di Markov di primo ordine:

$$P(w_1, w_2, \dots, w_n) = P(w_1) \times \prod_{i=2}^n P(w_i | w_{i-1})$$

La probabilità di ogni parola dipende solo dalla parola immediatamente precedente.

Un **modello trigram** (3-gram) corrisponde a una catena di Markov di secondo ordine:

$$P(w_1, w_2, \dots, w_n) = P(w_1) \times P(w_2 | w_1) \times \prod_{i=3}^n P(w_i | w_{i-2}, w_{i-1})$$

Ordini superiori catturano dipendenze a più lungo raggio ma richiedono stima di un numero esponenzialmente crescente di parametri.

Nel contesto dell'acquisizione linguistica infantile, i modelli n-gram forniscono un framework per comprendere come i bambini potrebbero tracciare e apprendere probabilità di transizione tra unità linguistiche. L'ipotesi è che attraverso l'esposizione ripetuta, i bambini

costruiscano implicitamente rappresentazioni delle frequenze di transizione, che poi guidano la segmentazione, il riconoscimento di pattern, e la generalizzazione.

Hidden Markov Models: Estendendo il Framework

Spesso nel linguaggio, le unità osservabili (suoni acustici, fonemi pronunciati) sono generate da stati sottostanti non direttamente osservabili (intenzioni fonologiche, categorie fonematiche astratte, stati articolatori). Gli Hidden Markov Models (HMM) estendono le catene di Markov per modellare tali situazioni.

Un HMM consiste di:

1. **Stati nascosti** $\{s_1, s_2, \dots, s_N\}$: Gli stati sottostanti del processo
2. **Osservazioni** $\{o_1, o_2, \dots, o_N\}$: I simboli osservabili emessi
3. **Matrice di transizione** A : Probabilità di transizione tra stati nascosti $A[i,j] = P(s_{t+1} = j \mid s_t = i)$
4. **Matrice di emissione** B : Probabilità di emettere osservazioni dagli stati $B[i,k] = P(o_t = k \mid s_t = i)$
5. **Distribuzione iniziale** π : Probabilità degli stati iniziali

La sequenza di osservazioni $O = o_1, o_2, \dots, o_N$ è generata da una sequenza nascosta di stati $S = s_1, s_2, \dots, s_N$ secondo il processo:

1. Inizia nello stato s_1 con probabilità $\pi[s_1]$
2. Emetti osservazione o_1 con probabilità $B[s_1, o_1]$
3. Transita allo stato s_2 con probabilità $A[s_1, s_2]$
4. Emetti osservazione o_2 con probabilità $B[s_2, o_2]$
5. E così via...

La probabilità di osservare una particolare sequenza O marginalizzando su tutte le possibili sequenze di stati nascosti è:

$$P(O) = \sum_S P(O, S) = \sum_S \pi[s_1] B[s_1, o_1] \prod_{t=1}^{T-1} A[s_t, s_{t+1}] B[s_{t+1}, o_{t+1}]$$

Tre problemi fondamentali sono associati agli HMM:

Problema di valutazione: Data una sequenza di osservazioni O e un modello $\lambda = (A, B, \pi)$, calcolare $P(O \mid \lambda)$. Questo viene risolto efficientemente usando l'algoritmo forward o backward, con complessità $O(N^2T)$ dove N è il numero di stati e T la lunghezza della sequenza.

Problema di decodifica: Data una sequenza O e un modello λ , trovare la sequenza di stati nascosti più probabile $S^* = \operatorname{argmax}_S P(S \mid O, \lambda)$. Questo viene risolto usando l'algoritmo di Viterbi, anch'esso con complessità $O(N^2T)$.

Problema di apprendimento: Dato un insieme di sequenze di osservazioni, trovare i parametri del modello $\lambda^* = \operatorname{argmax}_\lambda P(O \mid \lambda)$ che massimizzano la verosimiglianza dei dati. Questo viene risolto usando l'algoritmo di Baum-Welch (una forma di algoritmo Expectation-Maximization), che itera tra stimare gli stati nascosti dati i parametri correnti (E-step) e aggiornare i parametri date le stime degli stati (M-step).

Gli HMM sono stati applicati estensivamente nel riconoscimento automatico del parlato, dove gli stati nascosti corrispondono a fonemi o unità subfonemiche e le osservazioni sono feature acustiche estratte dal segnale audio. Nel contesto dell'acquisizione linguistica, gli HMM forniscono un modello di come i bambini potrebbero apprendere a mappare variabilità acustica continua su categorie fonematiche discrete, un processo cruciale per lo sviluppo fonologico.

La Sfida della Complessità e della Generalizzazione

Un problema centrale nell'applicazione di modelli markoviani al linguaggio è il trade-off tra potere espressivo e trattabilità computazionale. Modelli di ordine superiore (trigram, 4-gram, ecc.) possono catturare dipendenze a più lungo raggio e quindi fare predizioni più accurate, ma richiedono stima di un numero di parametri che cresce esponenzialmente con l'ordine.

Per un modello bigram su un vocabolario di dimensione V , il numero di parametri da stimare è $O(V^2)$. Per un trigram, è $O(V^3)$. Per linguaggi naturali con vocabolari di decine di migliaia di parole, questo diventa rapidamente intrattabile. Inoltre, anche con grandi corpora, molte combinazioni specifiche non saranno mai osservate, portando al problema dello **zero-probability**: assegnare probabilità zero a sequenze non viste ma potenzialmente valide.

Diverse tecniche sono state sviluppate per affrontare questi problemi:

Smoothing: Metodi come Laplace smoothing, Good-Turing smoothing, e Kneser-Ney smoothing redistribuiscono massa di probabilità dalle sequenze osservate a quelle non osservate, evitando probabilità zero ma preservando le frequenze relative apprese.

Back-off e interpolazione: Se una sequenza lunga (per esempio, un trigram) non è stata osservata, si "retrocede" a una sequenza più corta (bigram o unigram). L'interpolazione combina predizioni di modelli di diverso ordine con pesi appropriati.

Modelli gerarchici Bayesiani: Usando prior Bayesiani gerarchici, si può condividere forza statistica attraverso contesti simili, migliorando la generalizzazione anche con dati limitati.

Compressione dello spazio degli stati: Tecniche di clustering o embeddings possono ridurre la dimensionalità raggruppando stati simili.

Nel contesto dell'acquisizione linguistica infantile, questi problemi tecnici riflettono sfide cognitive reali che i bambini devono affrontare. L'input linguistico a cui sono esposti è finito e spesso contiene lacune. Devono generalizzare da esempi visti a casi nuovi in modi che sono produttivi ma non eccessivamente permissivi. I meccanismi cognitivi che i bambini usano per affrontare la sparsità dei dati e per generalizzare appropriatamente sono al centro della ricerca sull'acquisizione linguistica, e i modelli markoviani forniscono un framework computazionale per esplorare ipotesi specifiche su questi meccanismi.

Motivazioni Cognitive e Biologiche

Prima di procedere all'esplorazione dettagliata di come i modelli markoviani illuminano aspetti specifici dell'acquisizione linguistica, è utile considerare le motivazioni cognitive e biologiche per questo approccio.

L'apprendimento statistico sequenziale non è unico agli esseri umani o al linguaggio. Esperimenti con primati non umani, roditori, e persino uccelli hanno dimostrato sensibilità a probabilità di transizione in sequenze di stimoli. Questa ubiquità suggerisce che meccanismi neurali per tracciare statistiche temporali potrebbero essere evolutivamente antichi e dominio-generalisti.

A livello neurale, l'apprendimento di associazioni sequenziali è supportato da plasticità sinaptica dipendente dal tempo (spike-timing dependent plasticity, STDP) e da circuiti nei gangli della base che sono cruciali per l'apprendimento di sequenze motorie e cognitive. Studi di neuroimaging hanno identificato reti cerebrali coinvolte nell'apprendimento statistico implicito che si sovrappongono parzialmente ma non completamente con reti linguistiche classiche, supportando l'idea che meccanismi di apprendimento statistico dominio-generalisti contribuiscano all'acquisizione linguistica.

Comportamentalmente, i bambini dimostrano capacità di apprendimento statistico molto precocemente. Neonati di pochi giorni di vita discriminano tra lingue basandosi su pattern ritmici e prosodici. Bambini di otto mesi segmentano parole da flussi di parlato continuo basandosi su probabilità di transizione sillabica. Queste capacità emergono prima e in assenza di competenze linguistiche esplicite, suggerendo che rappresentano meccanismi fondazionali su cui si costruisce acquisizione linguistica più complessa.

L'integrazione di evidenze comportamentali, neurali, computazionali e comparative supporta un framework in cui modelli markoviani catturano aspetti genuini dei processi cognitivi sottostanti all'acquisizione linguistica infantile. Nelle sezioni successive, esamineremo in dettaglio come questi modelli applicano a domini specifici del linguaggio e come si confrontano con evidenze empiriche.

Parte II: Segmentazione del Parlato e Apprendimento Statistico delle Sequenze Sillabiche

Il Problema della Segmentazione: Dal Continuo al Discreto

Uno dei primi compiti fondamentali che un bambino deve affrontare nell'acquisizione linguistica è la segmentazione del parlato. A differenza del testo scritto, dove gli spazi tra parole forniscono delimitazioni esplicite, il parlato fluente è essenzialmente continuo. Non esistono pause acustiche affidabili tra parole; infatti, le pause che occorrono nel parlato naturale spesso cadono all'interno di frasi o gruppi prosodici piuttosto che tra parole. Il segnale acustico che raggiunge l'orecchio del bambino è un flusso ininterrotto di suoni, eppure il bambino deve segmentare questo flusso in unità discrete come parole, sillabe e fonemi.

Questo problema è stato quantificato sperimentalmente usando stimoli artificiali. Se si presenta a adulti parlanti di inglese una registrazione in italiano o mandarino, lingue che non conoscono, essi hanno grande difficoltà a identificare dove inizia e finisce una parola, anche se possono percepire chiaramente i suoni individuali. Questo dimostra che la segmentazione non è un processo puramente acustico-fonetico ma richiede conoscenza specifica della lingua.

Come fanno allora i bambini, che non hanno ancora acquisito tale conoscenza, a segmentare il parlato? La risposta tradizionale ha invocato vari indizi prosodici: pattern di accento, contorni intonazionali, allungamento preconfinale (l'ultima sillaba di una parola tende ad essere più lunga), e fenomeni fonetici specifici della lingua come l'aspirazione di consonanti in inglese che tende a occorrere all'inizio di parola. Sebbene questi indizi siano senza dubbio importanti, ricerche degli ultimi tre decenni hanno rivelato che i bambini sfruttano anche un meccanismo più generale e potente: l'apprendimento di probabilità di transizione tra unità sublessicali come sillabe.

Gli Esperimenti Seminali di Saffran, Aslin e Newport

Una serie di esperimenti eleganti condotti da Jenny Saffran, Richard Aslin e Elissa Newport alla fine degli anni Novanta ha rivoluzionato la nostra comprensione della segmentazione del parlato infantile. Nel loro paradigma sperimentale classico, bambini di otto mesi furono esposti a un flusso continuo di sillabe sintetizzate senza pause, per esempio:

bidakupadotigolabubidaku...

Questo flusso era costruito concatenando quattro "parole" artificiali trisillabiche:

bidaku, padoti, golabu, pidatu

Ogni "parola" era presentata molte volte (tipicamente l'esposizione totale era di circa due minuti), ma sempre nella stessa sequenza di tre sillabe. Crucialmente, non c'erano pause tra le parole, né altri indizi prosodici per i confini di parola. L'unico indizio disponibile era la struttura statistica: la probabilità di transizione tra sillabe all'interno di una parola era 1,0 (per esempio, dopo "bi" seguiva sempre "da", e dopo "da" seguiva sempre "ku"), mentre la probabilità di transizione tra l'ultima sillaba di una parola e la prima sillaba della parola successiva era molto più bassa (0,33 nell'esempio sopra, dato che tre diverse parole potevano seguire).

Dopo questa fase di familiarizzazione, i bambini furono testati usando il paradigma di preferenza uditiva (headturn preference procedure). Venivano presentate sia le "parole" familiari (per esempio, "bidaku") sia sequenze di controllo composte dalle stesse sillabe ma che attraversavano confini di parola durante l'addestramento (per esempio, "kupado", formato dalle ultime due sillabe di "bidaku" e la prima sillaba di "padoti"). Se i bambini avessero segmentato il flusso basandosi sulle probabilità di transizione, dovrebbero riconoscere le parole familiari come più coerenti o "simili a parola" rispetto ai controlli.

I risultati furono sorprendenti: i bambini di otto mesi ascoltarono significativamente più a lungo (mostrando maggiore interesse) alle "non-parole" rispetto alle parole familiari. Questo pattern di preferenza per la novità è interpretato come evidenza di riconoscimento: i bambini avevano codificato le parole familiari durante l'esposizione e quindi le trovavano meno interessanti delle sequenze nuove al test. Questo risultato dimostrò che bambini molto piccoli possono estrarre unità coerenti da un flusso continuo basandosi esclusivamente su statistiche di transizione, senza bisogno di indizi prosodici o semantici.

Formalizzazione Markoviana della Segmentazione

Possiamo formalizzare questo processo di segmentazione usando un modello markoviano. Consideriamo il flusso di parlato come una sequenza di sillabe $s_1, s_2, s_3, \dots, s_n$. Definiamo la probabilità di transizione tra sillabe:

$$P(s_{i+1} | s_i) = \text{count}(s_i, s_{i+1}) / \text{count}(s_i)$$

dove $\text{count}(s_i, s_{i+1})$ è il numero di volte che la sillaba s_{i+1} segue immediatamente s_i , e $\text{count}(s_i)$ è il numero totale di occorrenze di s_i .

Alternativamente, possiamo usare la **probabilità di transizione inversa**:

$$P(s_i | s_{i+1}) = \text{count}(s_i, s_{i+1}) / \text{count}(s_{i+1})$$

Una metrica particolarmente utile per identificare confini di parola è la **probabilità di transizione bidirezionale** o **mutua informazione puntuale** (pointwise mutual information):

$$MI(s_i, s_{i+1}) = \log[P(s_i, s_{i+1}) / (P(s_i) P(s_{i+1}))]$$

che misura quanto la co-occorrenza di s_i e s_{i+1} è più frequente di quanto ci si aspetterebbe se fossero indipendenti. Valori alti di MI indicano forte associazione (tipicamente all'interno di parole); valori bassi indicano associazione debole (tipicamente ai confini di parola).

Un'altra metrica rilevante è la **probabilità predittiva transizionale** (transitional probability, TP):

$$TP(s_i, s_{i+1}) = P(s_{i+1} | s_i)$$

Nel paradigma di Saffran e colleghi, le TP all'interno delle parole erano 1,0 mentre le TP ai confini erano 0,33. I bambini sembravano sensibili a questa differenza e la usavano per segmentare.

Un algoritmo semplice per la segmentazione basato su queste statistiche è:

1. Scandire la sequenza di sillabe da sinistra a destra
2. A ogni posizione i , calcolare $TP(s_i, s_{i+1})$
3. Se $TP(s_i, s_{i+1}) < \theta$ (una soglia), postulare un confine di parola tra s_i e s_{i+1}
4. Altrimenti, considerare s_i e s_{i+1} come appartenenti alla stessa parola

La scelta della soglia θ è critica. Una soglia troppo alta porterà a sotto-segmentazione (poche parole, troppo lunghe); una soglia troppo bassa porterà a sovra-segmentazione (molte parole, troppo corte). Studi hanno esplorato come i bambini potrebbero adattivamente impostare questa soglia, possibilmente basandosi sulla distribuzione complessiva delle TP osservate.

Modelli Bayesiani di Segmentazione

Approcci Bayesiani forniscono un framework più sofisticato per la segmentazione che non richiede impostazione manuale di soglie ma invece bilancia esplicitamente diversi fattori: adattarsi ai dati osservati, preferire soluzioni parsimoniose, e incorporare bias a priori.

Nel modello Bayesiano di segmentazione di parole, l'obiettivo è trovare la segmentazione ottimale S^* data una sequenza di sillabe osservata x :

$$S^* = \operatorname{argmax}_S P(S | x) = \operatorname{argmax}_S P(x | S) P(S)$$

dove $P(x | S)$ è la verosimiglianza (quanto bene la segmentazione spiega i dati) e $P(S)$ è il prior (quanto la segmentazione è intrinsecamente plausibile).

Un prior comunemente usato favorisce lessici più piccoli e parole più brevi:

$$P(S) \propto \exp(-\alpha |\operatorname{lexicon}(S)| - \beta \sum_w |w|)$$

dove $\operatorname{lexicon}(S)$ è l'insieme delle parole distinte nella segmentazione, $|w|$ è la lunghezza della parola w , e α , β sono parametri che controllano la forza di queste preferenze.

La verosimiglianza può essere modellata assumendo che le parole siano generate indipendentemente secondo un modello unigram:

$$P(x | S) = \prod_w P(w)^{\operatorname{count}(w)}$$

dove il prodotto è su tutte le occorrenze di parole nella segmentazione.

Per segmentare, si usa tipicamente campionamento di Gibbs: si inizia con una segmentazione casuale, poi si itera scandendo ogni possibile confine di parola e decidendo (stocasticamente) se inserire o rimuovere un confine basandosi sulla probabilità posteriore condizionale.

Goldwater, Griffiths e Johnson (2009) hanno dimostrato che tali modelli Bayesiani, quando addestrati su trascrizioni fonetiche di parlato infantile-diretto (child-directed speech), possono segmentare parole con accuratezza sorprendente (precision ~75%, recall ~70%), ben superiore a algoritmi baseline puramente statistici. Questo suggerisce che combinare sensibilità statistica con bias appropriati può portare a segmentazione efficace anche senza supervisione esplicita.

Estensioni: Ordine Superiore e Contesto

I modelli bigram (Markov di primo ordine) catturano dipendenze tra coppie adiacenti di sillabe. Tuttavia, evidenze sperimentali suggeriscono che i bambini potrebbero tracciare anche dipendenze di ordine superiore.

Un modello **trigram** condiziona su due sillabe precedenti:

$$P(s_{i+1} | s_i, s_{i-1})$$

Questo può catturare pattern come "dopo la sequenza 'bi-da' segue quasi sempre 'ku'", che fornisce informazioni più ricche per segmentazione.

Esperimenti con adulti e bambini hanno mostrato sensibilità a TP di ordine superiore. Per esempio, Pelucchi, Hay e Saffran (2009) dimostrarono che bambini di otto mesi possono tracciare probabilità di transizione non solo tra sillabe adiacenti ma anche tra sillabe

separate da una sillaba intermedia (dipendenze non-adiacenti). Questo suggerisce che il sistema di apprendimento statistico infantile è più sofisticato di un semplice contatore bigram.

Formalmente, possiamo estendere il modello Markoviano per includere dipendenze non-adiacenti usando **skip-bigrams** o **sparse Markov models** dove le transizioni possono "saltare" elementi intermedi:

$$P(s_{i+2} | s_i) \text{ oppure } P(s_{i+3} | s_i)$$

Alternativamente, possiamo usare **Variable-order Markov Models** (VMM) o **Prediction Suffix Trees** che adattano dinamicamente l'ordine del modello: usano contesto lungo quando informativo e contesto breve quando il contesto lungo non aggiunge informazione predittiva.

Integrazione di Indizi Multipli: Modelli Multifattoriali

Nella realtà, i bambini non si basano esclusivamente su statistiche di transizione ma integrano molteplici fonti di informazione: prosodia, fonologia, ritmo, e gradualmente anche informazione semantica e sintattica.

Modelli computazionali sofisticati hanno esplorato come diversi indizi possano essere combinati. Per esempio:

Indizi prosodici: Confini di parola tendono a coincidere con confini di gruppi prosodici o con cambiamenti nell'altezza tonale (pitch). Questi possono essere modellati come feature aggiuntive che modulano le probabilità di confine.

Fonotassi: Alcune sequenze di suoni sono permesse all'interno di parole in una lingua ma non ai confini. Per esempio, in inglese, "kt" può occorrere ai confini di parola ("baked tomatoes") ma raramente all'interno. Vincoli fonotattici possono essere codificati come penalità o bonus sulle probabilità di segmentazione.

Armonia vocalica: In lingue come turco o ungherese, le vocali all'interno di parole tendono a condividere certi tratti. Questo fornisce un indizio aggiuntivo per coesione intra-parola.

Un modello integrato assegnerebbe probabilità di confine come funzione di molteplici features:

$$P(\text{boundary at position } i) = \sigma(w_1 \times \text{TP_feature} + w_2 \times \text{prosody_feature} + w_3 \times \text{phonotactic_feature} + \dots)$$

dove σ è una funzione sigmoide e i pesi w possono essere appresi da dati.

Johnson e collaboratori hanno sviluppato modelli che combinano TP con indizi prosodici e hanno mostrato che la combinazione supera significativamente ciascun indizio da solo, raggiungendo accuratèzze superiori all'80% su corpora di parlato infantile-diretto reale.

Evidenze Sperimentali Oltre il Laboratorio

Mentre gli esperimenti di Saffran con lingue artificiali hanno fornito evidenza causale che i bambini possono usare statistiche di transizione per segmentazione, sorge la domanda: usano effettivamente questo meccanismo nell'acquisizione del linguaggio naturale?

Studi correlazionali: Ricerche hanno analizzato le statistiche di transizione nel parlato infantile-diretto reale e hanno trovato che i confini di parola effettivi tendono a corrispondere a posizioni dove le TP sono relativamente basse. Questo suggerisce che la struttura statistica del parlato naturale fornisce effettivamente indizi utili per segmentazione.

Studi longitudinali: Seguendo bambini nel tempo, ricercatori hanno trovato correlazioni tra la sensibilità a pattern statistici misurata in laboratorio a un'età precoce e il successivo sviluppo del vocabolario. Bambini che mostrano maggiore sensibilità statistica a 8 mesi tendono ad avere vocabolari più ampi a 18-24 mesi. Questo supporta l'idea che apprendimento statistico gioca un ruolo causale nell'acquisizione reale.

Variazioni cross-linguistiche: Lingue diverse hanno strutture statistiche diverse. Per esempio, lingue sillabo-temporizzate come italiano o spagnolo hanno sillabe di durata relativamente uniforme, mentre lingue accento-temporizzate come inglese hanno maggiore variabilità. Studi hanno mostrato che bambini esposti a lingue diverse mostrano sensibilità a pattern statistici specifici della loro lingua già nei primi mesi di vita.

Limiti e Critiche dei Modelli Markoviani per Segmentazione

Nonostante i successi, i modelli markoviani per segmentazione hanno limiti riconosciuti:

Problema del bootstrapping: Per calcolare TP tra sillabe, il bambino deve già aver segmentato il flusso acustico in sillabe. Ma la segmentazione sillabica stessa è non-triviale. Come si rompe questa circolarità? Possibilmente attraverso proprietà acustiche universali delle sillabe (alternanza di segmenti ad alta e bassa sonorità) o attraverso co-sviluppo iterativo di segmentazione a molteplici livelli.

Sottosegmentazione e sovrasegmentazione: Modelli puramente statistici tendono a fare errori sistematici, segmentando frequenti frasi fisse come unità singole (per esempio, "dov'è" potrebbe non essere segmentato) o sovra-segmentando parole lunghe rare. Bambini reali fanno errori simili nelle prime fasi ma li correggono gradualmente, suggerendo meccanismi aggiuntivi di revisione e raffinamento.

Mancanza di generalizzazione astratta: I modelli Markoviani traccianti TP specifiche tra token specifici non catturano generalizzazioni a livelli più astratti (per esempio, pattern prosodici che dipendono da categorie di stress piuttosto che sillabe specifiche). Questo richiede rappresentazioni gerarchiche più sofisticate.

Computazione e memoria: Tracciare tutte le TP possibili tra tutte le coppie di sillabe richiede memoria considerevole. Meccanismi cognitivi plausibili potrebbero dover usare approssimazioni, dimenticanza di informazioni vecchie, o rappresentazioni compresse.

Questi limiti hanno motivato lo sviluppo di modelli più sofisticati che combinano apprendimento statistico con rappresentazioni strutturate, bias linguistici specifici, e meccanismi di attenzione e memoria realistici. Tuttavia, il framework markoviano

fondamentale rimane centrale, fornendo la base computazionale su cui si costruiscono questi meccanismi aggiuntivi.

Meccanismi Neurali dell'Apprendimento Statistico

Domanda cruciale: quali sono i meccanismi cerebrali che implementano questo tracciamento di probabilità di transizione?

Studi di neuroimaging con adulti hanno identificato una rete distribuita coinvolta nell'apprendimento statistico, includendo:

Aree temporali superiori: Corteccia uditiva primaria e associativa, coinvolte nell'elaborazione di suoni e sequenze auditive.

Ippocampo: Cruciale per apprendimento rapido di associazioni e per codifica di sequenze temporali.

Gangli della base: Particolarmente nucleo caudato e striatum, noti per ruolo nell'apprendimento procedurale e nella predizione.

Corteccia prefrontale: Coinvolta in mantenimento di rappresentazioni in memoria di lavoro e in integrazione di informazioni nel tempo.

Studi con ERP (event-related potentials) hanno mostrato che violazioni di pattern statistici appresi elicitano risposte neurali specifiche, come componenti N1 e P300 aumentate, suggerendo che il cervello forma automaticamente rappresentazioni predittive e rileva deviazioni da queste predizioni.

A livello di circuiti neurali, modelli computazionali hanno proposto che l'apprendimento di sequenze statistiche possa essere implementato attraverso plasticità hebbiana nelle connessioni sinaptiche: sinapsi tra neuroni che sono ripetutamente attivati in sequenza si rafforzano, codificando implicitamente le probabilità di transizione. Questo fornisce un meccanismo biologicamente plausibile per apprendimento markoviano.

Apprendimento Statistico come Meccanismo Dominio-Generale

Un dibattito importante riguarda se l'apprendimento statistico che supporta segmentazione del parlato sia specificamente linguistico o rappresenti una capacità cognitiva dominio-generale.

Evidenze per dominio-generalità:

Apprendimento visuo-spaziale: Adulti e bambini mostrano sensibilità a pattern statistici in sequenze visive (forme che appaiono in ordini specifici) con performance simile a compiti uditivo-linguistici.

Apprendimento motorio: Pattern di movimenti sequenziali possono essere appresi attraverso esposizione ripetuta a regolarità statistiche.

Apprendimento in altre specie: Primati non umani, roditori e uccelli mostrano capacità di apprendimento statistico in vari domini.

Evidenze per specificità o modulazione:

Vantaggi per stimoli linguistici: Alcuni studi trovano che l'apprendimento è più robusto per stimoli linguistici rispetto a non-linguistici, suggerendo specializzazione.

Correlazioni selettive: Performance in compiti di apprendimento statistico con stimoli linguistici predice abilità linguistiche meglio di quanto faccia performance con stimoli non-linguistici.

Sviluppo: La traiettoria di sviluppo dell'apprendimento statistico può differire tra domini.

La visione più plausibile sembra essere che esista un meccanismo computazionale di base dominio-generale per tracciare regolarità statistiche, ma che questo meccanismo è modulato e ottimizzato attraverso sviluppo per elaborazione di input linguistici, riflettendo sia capacità innate dominio-generalis sia apprendimento specifico al dominio attraverso esperienza.

La segmentazione del parlato, vista attraverso la lente dei modelli markoviani e dell'apprendimento statistico, emerge come un processo multifattoriale che sfrutta sia proprietà universali della computazione neurale (capacità di tracciare co-occorrenze e predire eventi futuri basati su esperienze passate) sia strutture specifiche del segnale linguistico. Questo processo rappresenta il primo passo critico nell'acquisizione linguistica, trasformando un flusso continuo e ambiguo in unità discrete che possono poi essere mappate su significati e integrate in sistemi grammaticali più complessi.

Parte III: Acquisizione Fonologica e Apprendimento dei Pattern Fonotattici

La Costruzione dello Spazio Fonologico

Una volta che i bambini iniziano a segmentare il parlato in unità più piccole, devono affrontare il compito ancora più fondamentale di identificare le unità di suono distintive della loro lingua: i fonemi. I fonemi sono le unità sonore minimali che distinguono significati. Per esempio, in italiano, /p/ e /b/ sono fonemi distinti perché cambiano il significato ("pane" vs "bane"), mentre le variazioni acustiche nella pronuncia esatta di /p/ da parte di diversi parlanti non sono distintive fonologicamente.

Il problema centrale dell'acquisizione fonologica è che lo spazio acustico è continuo e ad alta dimensionalità, mentre le categorie fonemiche sono discrete e specifiche della lingua. Ogni lingua del mondo seleziona un sottoinsieme diverso delle possibili distinzioni fonetiche come fonologicamente rilevanti. Per esempio, in mandarino, diverse altezze tonali (toni) sono fonologicamente contrastive e cambiano il significato delle parole, mentre in italiano i toni non hanno questa funzione. In inglese, l'aspirazione della /p/ iniziale in "pin" vs la mancanza di aspirazione in "spin" è una variazione allofonica non distintiva, mentre in thai questa stessa distinzione è fonologica.

I neonati nascono come "cittadini del mondo fonetico": possono discriminare contrasti fonetici di tutte le lingue del mondo. Studi classici di Patricia Kuhl, Janet Werker e collaboratori hanno dimostrato che neonati di 6-8 mesi discriminano contrasti fonetici non presenti nella loro lingua nativa (per esempio, bambini giapponesi discriminano /r/ da //, contrasto che non esiste in giapponese). Tuttavia, entro 10-12 mesi, questa sensibilità universale si riduce: i bambini diventano meno capaci di discriminare contrasti non presenti nella loro lingua, mentre migliorano nella discriminazione dei contrasti nativi. Questo processo è stato chiamato "perceptual narrowing" o "neural commitment".

Come avviene questa riorganizzazione percettiva? I modelli markoviani forniscono un framework per comprendere come l'esposizione statistica ai pattern sonori della lingua nativa guida la formazione di categorie fonemiche.

Modelli di Acquisizione Fonologica Basati su Distribuzione

Un approccio influente è basato sull'ipotesi che i bambini usano informazioni sulla distribuzione delle varianti fonetiche nell'input per inferire le categorie fonemiche sottostanti. Questa prospettiva è radicata in idee strutturaliste classiche di Roman Jakobson e Nikolai Trubetzkoy sulla fonologia come sistema di opposizioni distintive.

L'idea chiave è che allofonini dello stesso fonema (varianti fonetiche che non cambiano il significato) tendono a occorrere in **distribuzione complementare**: ciascun allofono appare in contesti specifici e non si sovrappongono. Per esempio, in inglese, [p^h] (p aspirata) appare tipicamente all'inizio di sillaba tonica, mentre [p] (p non aspirata) appare dopo /s/ o in fine di sillaba. Non troviamo coppie minime dove [p^h] vs [p] cambia il significato. Al contrario, fonemi distinti appaiono in **distribuzione contrastiva**: possono occorrere negli stessi contesti e cambiano il significato.

Un modello markoviano può catturare questo attraverso probabilità di transizione condizionate da contesto. Definiamo uno stato come coppia (suono, contesto):

$P(\text{suono}_i | \text{contesto}_i)$

Se due suoni hanno distribuzioni molto diverse condizionate su contesti (uno appare prevalentemente in contesti dove l'altro è raro), questo suggerisce che sono allofoni dello stesso fonema. Se due suoni appaiono negli stessi contesti con frequenze simili, questo suggerisce che sono fonemi distinti.

Formalmente, possiamo usare **divergenza di Kullback-Leibler** per misurare la differenza tra distribuzioni:

$$KL(P(\text{context} | \text{sound}_1) || P(\text{context} | \text{sound}_2)) = \sum_c P(c|s_1) \log[P(c|s_1) / P(c|s_2)]$$

Valori bassi di KL suggeriscono distribuzione simile (fonemi distinti); valori alti suggeriscono distribuzione complementare (allofoni).

Maye, Werker e Gerken (2002) testarono una predizione di questo approccio distribuzionale. Esposero bambini di 6 e 8 mesi a distribuzioni artificiali di suoni lungo un continuum acustico. In una condizione, i suoni avevano distribuzione bimodale (due picchi separati), suggerendo due categorie distinte. In un'altra condizione, avevano distribuzione

unimodale (un solo picco), suggerendo una singola categoria con variazione. Dopo esposizione, i bambini nella condizione bimodale discriminarono meglio i suoni ai due estremi del continuum, mentre quelli nella condizione unimodale trattarono tutti i suoni come varianti della stessa categoria. Questo dimostrò che bambini molto piccoli usano informazioni distribuzionali per formare categorie fonologiche.

Hidden Markov Models per Fonologia

Gli Hidden Markov Models (HMM) forniscono un framework naturale per modellare acquisizione fonologica perché catturano esplicitamente la distinzione tra rappresentazioni fonologiche astratte (stati nascosti) e realizzazioni fonetiche variabili (osservazioni).

In questo modello, gli stati nascosti corrispondono a fonemi o categorie fonemiche, e le osservazioni corrispondono a realizzazioni acustiche specifiche. La matrice di emissione $B[\text{fonema}, \text{suono_acustico}]$ specifica la probabilità che un dato fonema sia realizzato come un particolare suono acustico in un dato contesto.

Per esempio, per il fonema /p/ in inglese:

$B[p/, [p^h]]$ = alta probabilità in posizione iniziale tonica $B[p/, [p]]$ = alta probabilità dopo /s/ o in coda $B[p/, [\bar{p}]]$ = alta probabilità in finale assoluto (release non udibile)

Il compito del bambino è inferire questa struttura latente (quali fonemi esistono e come vengono realizzati) dall'input acustico osservato. Questo è esattamente il problema di apprendimento in HMM.

Vallabha, McClelland, Pons, Werker e Amano (2007) hanno sviluppato un modello computazionale basato su HMM dell'acquisizione della distinzione /r/-/l/ in inglese. Il loro modello prende come input distribuzioni di feature acustiche estratte da registrazioni di parlato e usa algoritmi di clustering per identificare categorie fonemiche. Il modello simula con successo sia l'iniziale sensibilità universale (perché inizialmente può formare qualsiasi categoria presente nell'input) sia il successivo narrowing percettivo (perché le categorie si stabilizzano sui fonemi effettivamente presenti nella lingua target, e la discriminazione di contrasti assenti diventa più difficile).

Fonotassi: Vincoli su Sequenze Permesse

Oltre all'inventario fonetico, ogni lingua ha regole fonotattiche che specificano quali sequenze di suoni sono permesse e quali sono proibite. Per esempio, in italiano, una parola può iniziare con "str-" ("strada"), ma non con "tl-" (nessuna parola italiana inizia così, sebbene questa sequenza esista in lingue come nahuatl). Alcune di queste restrizioni sono universali (riflessioni di vincoli articolatori), altre sono specifiche della lingua.

I bambini devono apprendere i vincoli fonotattici della loro lingua. Questo apprendimento può essere modellato come acquisizione di probabilità di transizione tra fonemi in un modello markoviano.

Definiamo un modello bigram sui fonemi:

$$P(\text{fonema}_{i+1} | \text{fonema}_i)$$

Alcune transizioni avranno probabilità alta (sequenze comuni nella lingua), altre probabilità bassa o zero (sequenze rare o impossibili).

Un modello più sofisticato condiziona su posizione nella parola:

$$P(\text{fonema}_i \mid \text{fonema}_{i-1}, \text{posizione}_i)$$

distinguendo tra inizio di parola, interno di parola, fine di parola, poiché vincoli fonotattici differiscono secondo posizione.

Jusczyk, Friederici, Wessels, Svenkerud e Jusczyk (1993) dimostrarono che bambini di 9 mesi preferiscono ascoltare parole che rispettano la fonotassi della loro lingua nativa rispetto a parole che la violano, anche quando tutte le parole sono composte di fonemi nativi. Per esempio, bambini inglesi preferirono "chun" e "yush" (che rispettano pattern fonotattici inglesi) a "hnuv" e "zlor" (che li violano). Questo dimostrò che già a 9 mesi i bambini hanno estratto regolarità statistiche sui pattern di suoni permessi.

Vincoli di Sonorità e Gerarchia Prosodica

Molte restrizioni fonotattiche si basano sulla nozione di **sonorità** (sonority), una scala di prominenza acustica dove vocali hanno sonorità massima, approssimanti e liquide hanno sonorità intermedia, nasali sono meno sonore, fricative ancora meno, e occlusive hanno sonorità minima.

Un principio quasi-universale è il **Principio di Sequenziamento della Sonorità** (Sonority Sequencing Principle): all'interno di sillabe, la sonorità tende a salire verso il nucleo (tipicamente una vocale) e a scendere dopo. Questo spiega perché "pla" è una sillaba possibile in molte lingue (sonorità crescente $p < l < a$), mentre "lpa" è molto raro (sonorità non monotona).

I bambini devono apprendere queste tendenze universali e anche eccezioni e pattern specifici della loro lingua. Un modello markoviano può codificare il principio di sonorità attraverso probabilità di transizione che favoriscono sequenze conformi:

$$P(f_{i+1} \mid f_i, \text{position}) = f(\text{sonority_profile}(f_i \rightarrow f_{i+1}))$$

dove la probabilità è modulata da quanto la transizione rispetta gradienti di sonorità attesi.

Studi computazionali hanno mostrato che modelli che incorporano bias per principi di sonorità universali apprendono vincoli fonotattici più rapidamente ed efficacemente rispetto a modelli che trattano tutte le sequenze ugualmente a priori. Questo suggerisce che l'acquisizione fonologica beneficia di interazione tra apprendimento statistico dominio-generale e bias percettivi o articolatori innati.

Apprendimento di Processi Fonologici

Le lingue non solo hanno vincoli su sequenze permesse, ma anche **processi fonologici** che trasformano rappresentazioni fonologiche sottostanti in forme fonetiche superficiali. Per esempio, in italiano, la nasale /n/ assimila in punto di articolazione alla consonante seguente:

/n/ + /p/ → [m]p (come in "in pace" → [im'patʃe]) /n/ + /k/ → [ŋ]k (come in "in casa" → [in'kasa])

I bambini devono apprendere questi pattern di alternanza. Un modello markoviano può rappresentare questo attraverso probabilità condizionate:

$P(\text{suono_superficiale} \mid \text{suono_sottostante}, \text{contesto})$

L'apprendimento consiste nell'inferire queste mappature dall'osservazione di alternanze nell'input.

Hayes e Wilson (2008) hanno sviluppato un modello computazionale che apprende vincoli fonotattici e processi fonologici usando maximum entropy (modello log-lineare). Il loro modello assegna probabilità a sequenze fonologiche come funzione di features locali:

$P(\text{sequenza}) \propto \exp(\sum_i w_i f_i(\text{sequenza}))$

dove f_i sono features (presenza di certe sequenze, violazioni di certi vincoli) e w_i sono pesi appresi da dati. Quando addestrato su lessici di varie lingue, il modello riproduce accuratamente pattern fonotattici osservati e genera "parole possibili" che sono giudicate come fonotaticamente ben formate da parlanti nativi.

Modelli Bayesiani Non Parametrici per Inventari Fonemici

Una sfida per modelli di acquisizione fonologica è che il numero di fonemi non è noto a priori e varia da lingua a lingua (da 11 fonemi in rotokas a oltre 100 in !xóõ). Il bambino deve quindi non solo apprendere quali sono i fonemi, ma anche quanti sono.

Modelli Bayesiani non parametrici, particolarmente il **Dirichlet Process** e modelli di mixture come il **Infinite Gaussian Mixture Model**, forniscono soluzioni eleganti. Questi modelli permettono al numero di cluster (categorie fonemiche) di essere inferito dai dati piuttosto che specificato a priori.

Nel Dirichlet Process Mixture Model per fonologia:

- Ogni osservazione acustica (frame acustico o vettore di features) è generata da una categoria fonologica (cluster)
- Il numero di categorie è potenzialmente infinito ma determinato dai dati
- Nuove osservazioni possono essere assegnate a cluster esistenti o creare nuovi cluster
- La probabilità di assegnare un'osservazione a un cluster dipende da quante osservazioni precedenti sono state assegnate a quel cluster (effetto "rich get richer") e da quanto l'osservazione è simile alle osservazioni già nel cluster

Formalmente, se z indica assegnazioni di osservazioni a cluster:

$P(z_n = k \mid z_1, \dots, z_{n-1}, \alpha) \propto \{ n_k / (n - 1 + \alpha) \text{ se } k \text{ è un cluster esistente (con } n_k \text{ membri)} \alpha / (n - 1 + \alpha) \text{ se } k \text{ è un nuovo cluster} \}$

dove α è un parametro di concentrazione che controlla la tendenza a creare nuovi cluster.

Feldman, Griffiths e Morgan (2009) hanno applicato questo framework all'acquisizione di categorie fonemiche, mostrando che il modello può segmentare spazio acustico continuo in categorie discrete in modo che corrisponde qualitativamente allo sviluppo osservato nei bambini, incluso il narrowing percettivo.

Interazione tra Livelli: Lessico e Fonologia

Un insight importante è che acquisizione fonologica non avviene in isolamento ma interagisce con sviluppo lessicale. Mentre i bambini apprendono parole, usano la conoscenza delle parole per raffinare categorie fonemiche, e usano categorie fonemiche per facilitare apprendimento di nuove parole.

Questo porta a modelli di **bootstrapping reciproco** dove inferenza su fonologia e lessico si informano mutuamente. In un modello integrato:

- Le parole sono sequenze di fonemi: parola = fonema₁, fonema₂, ...
- I fonemi sono realizzati con variabilità acustica: fonema → distribuzione su segnali acustici
- Dati segnali acustici, inferire simultaneamente: (1) segmentazione in parole, (2) identità fonematica di segmenti

Questo può essere formulato come inferenza congiunta in un HMM gerarchico dove il livello superiore è un modello di sequenze di parole, il livello medio è un modello di sequenze di fonemi all'interno di parole, e il livello inferiore è un modello di realizzazioni acustiche di fonemi.

Studi computazionali hanno dimostrato che questo apprendimento congiunto supera apprendimento di ciascun livello separatamente, perché informazioni da un livello vincolano e guidano inferenza nell'altro.

Differenze Individuali e Traiettorie di Sviluppo

Non tutti i bambini seguono esattamente la stessa traiettoria nell'acquisizione fonologica. Alcuni bambini mostrano patterns di errore caratteristici (per esempio, sostituzioni sistematiche di certi suoni), mentre altri mostrano patterns diversi. Queste differenze individuali possono riflettere variazioni nei meccanismi di apprendimento statistico o nella qualità/quantità dell'input.

Studi recenti hanno iniziato a modellare queste differenze usando approcci computazionali. Per esempio, variando parametri in modelli Bayesiani (come la forza del prior o il tasso di apprendimento) si possono simulare diverse traiettorie di sviluppo che corrispondono a variazione osservata nella popolazione infantile.

Inoltre, evidenze crescenti suggeriscono che variabilità nell'input (per esempio, esposizione a molteplici dialetti o parlanti) può effettivamente facilitare certe forme di apprendimento fonologico, forzando il bambino a formare rappresentazioni più astratte e robuste piuttosto che memorizzare dettagli idiosincratici di singoli parlanti.

Plasticità e Periodi Critici

Una questione centrale riguarda la plasticità dell'acquisizione fonologica. Sebbene il narrowing percettivo avvenga entro il primo anno, i sistemi fonologici continuano a svilupparsi e raffinarsi nell'infanzia. Contrasti fonologici particolarmente difficili possono non essere completamente padroneggiati fino ai 7-8 anni.

Inoltre, esiste dibattito sull'esistenza di un "periodo critico" per acquisizione fonologica. L'evidenza più chiara per un periodo critico viene da studi su acquisizione di seconda lingua: parlanti che iniziano ad apprendere una seconda lingua dopo pubertà raramente raggiungono pronuncia nativa, mentre apprendenti precoci spesso lo fanno. Questo suggerisce che plasticità fonologica diminuisce con età.

Dal punto di vista dei modelli markoviani, il narrowing percettivo e la riduzione della plasticità possono essere comprese come conseguenze dell'apprendimento stesso: una volta che categorie si sono formate e consolidate, diventa progressivamente più difficile ristrutturarle. Questo è analogo al fenomeno di "overtraining" in reti neurali dove pattern appresi precocemente diventano radicati e resistenti a modificazione.

Modelli Bayesiani catturano questo attraverso la forza crescente dei prior: le prime esperienze stabiliscono distribuzioni prior che poi influenzano fortemente interpretazioni di esperienze successive. Con esperienza cumulativa, i prior diventano sempre più informativi (credenze concentrate) e quindi nuove evidenze hanno impatto minore.

Acquisizione Fonologica in Popolazioni Atipiche

Studiare acquisizione fonologica in popolazioni con sviluppo atipico fornisce insights su meccanismi sottostanti. Bambini con disturbi specifici del linguaggio (SLI), dislessia, o sordità congenita con impianti cocleari mostrano patterns diversi di sviluppo fonologico.

SLI: Questi bambini spesso mostrano persistenti difficoltà con discriminazioni fonemiche fini e con apprendimento di pattern fonotattici complessi. Modelli computazionali suggeriscono che ridotta capacità di tracciare statistiche temporali rapide potrebbe contribuire.

Dislessia: Bambini con dislessia spesso hanno difficoltà con consapevolezza fonologica (capacità di manipolare esplicitamente suoni in parole). Modelli suggeriscono che potrebbe essere dovuto a rappresentazioni fonologiche meno precise o "noisy" piuttosto che a deficit in apprendimento statistico per sé.

Impianti cocleari: Bambini sordi che ricevono impianti cocleari mostrano traiettorie variabili di sviluppo fonologico, dipendenti da età di impianto. Impianti precoci (primo anno) portano a sviluppo più vicino a normativo, coerente con idee di periodi critici.

Modellare queste popolazioni aiuta a dissociare componenti del sistema di acquisizione e a testare predizioni specifiche su quali meccanismi sono necessari per sviluppo fonologico normale.

Conclusioni sulla Fonologia come Sistema Markoviano

L'acquisizione fonologica, vista attraverso modelli markoviani e di apprendimento statistico, emerge come un processo in cui esposizione cumulativa a pattern sonori della lingua porta

gradualmente a formazione di categorie discrete e di rappresentazioni delle regolarità fonotattiche. Questo processo:

- Inizia con sensibilità universale a contrasti fonetici, supportata da predisposizioni percettive
- Progressivamente si specializza attraverso apprendimento statistico delle distribuzioni e transizioni nell'input nativo
- Coinvolge formazione di rappresentazioni astratte (fonemi) da variabilità acustica continua
- Interagisce con sviluppo lessicale e altri aspetti del linguaggio in modo reciproco
- È modulato da differenze individuali, qualità dell'input, e vincoli neurobiologici

I modelli markoviani, particolarmente HMM e modelli Bayesiani, forniscono framework computazionali precisi per formalizzare e testare queste idee, collegando teoria linguistica, evidenze comportamentali, e meccanismi neurali in una prospettiva integrata.

Parte IV: Acquisizione Morfologica e Regolarità Inflessionali

La Struttura Interna delle Parole

Dopo aver segmentato il parlato in parole e appreso il sistema fonologico, i bambini devono affrontare la scoperta della struttura interna delle parole. La morfologia studia come le parole sono costruite da unità più piccole di significato chiamate morfemi. Un morfema è l'unità minimale di significato o funzione grammaticale. Per esempio, la parola italiana "gattini" contiene tre morfemi: "gatt-" (radice nominale con significato lessicale 'gatto'), "-in-" (suffisso diminutivo), e "-i" (desinenza del plurale).

Le lingue variano enormemente nella complessità morfologica. L'inglese e il cinese mandarino sono morfologicamente relativamente semplici (lingue isolanti o analitiche), con poca inflazione e derivazione. Il latino, il russo e l'italiano sono lingue fusionali con sistemi inflessionali ricchi. Il turco, il giapponese e molte lingue amerindie sono agglutinant, con lunghe catene di morfemi che si concatenano per esprimere molteplici distinzioni grammaticali. Le lingue polisintetiche come l'inuktitut hanno parole che equivalgono a intere frasi in lingue analitiche.

L'acquisizione morfologica presenta sfide specifiche:

1. **Identificazione dei confini morfemici:** Dove finisce una radice e inizia un affisso?
2. **Categorizzazione:** Quali varianti fonetiche rappresentano lo stesso morfema (allomorfia)?
3. **Produttività:** Quali pattern sono regolari e generalizzabili vs idiosincratici?
4. **Astrazione:** Come si estraggono schemi astratti da esempi concreti?

I modelli markoviani forniscono strumenti per affrontare ciascuna di queste sfide.

Segmentazione Morfologica come Problema Markoviano

Analogamente alla segmentazione di parole dal parlato continuo, la segmentazione morfologica può essere formulata come problema di trovare confini ottimali all'interno di

parole. Se consideriamo una parola come sequenza di caratteri o fonemi $c_1c_2\dots c_n$, dobbiamo decidere dove inserire confini per dividere la parola in morfemi.

Un approccio basato su modelli markoviani usa **probabilità di transizione tra caratteri** come indizio per confini morfemici. L'intuizione è che transizioni tra caratteri all'interno di morfemi hanno probabilità diverse rispetto a transizioni ai confini morfemici.

Definiamo la probabilità di transizione:

$$P(c_{i+1} \mid c_i, \text{posizione})$$

dove "posizione" indica se siamo all'interno di morfema, a inizio di morfema, o a fine di morfema.

Un algoritmo di segmentazione morfologica calcola per ogni posizione potenziale di confine una misura di "plausibilità del confine" basata su calo della probabilità di transizione:

$$\text{boundary_score}(i) = -\log P(c_{i+1} \mid c_i)$$

Posizioni con punteggi alti sono candidati per confini morfemici.

Tuttavia, questo approccio puramente locale ha limiti. Un approccio più sofisticato usa **modelli di lingua a livello di morfema** che bilanciano coesione interna dei morfemi e plausibilità del lessico morfologico risultante.

Modelli Bayesiani di Segmentazione Morfologica

Goldsmith (2001) ha sviluppato un influente modello Bayesiano per apprendimento morfologico non supervisionato chiamato Linguistica. Il modello cerca di trovare una segmentazione del lessico in morfemi che massimizza una funzione obiettivo bayesiana:

$$\text{Score}(\text{Segmentation}) = \log P(\text{Data} \mid \text{Segmentation}) + \log P(\text{Segmentation})$$

Il termine di verosimiglianza favorisce segmentazioni che spiegano bene i dati osservati (parole nel corpus). Il termine di prior favorisce segmentazioni parsimoniose con meno morfemi distinti.

Il prior può essere formulato usando Minimum Description Length (MDL):

$$P(\text{Segmentation}) \propto \exp(-[\text{length}(\text{morpheme_lexicon}) + \text{length}(\text{morpheme_combinations})])$$

dove "length" misura la complessità di codifica.

Questo approccio è stato esteso in vari modi:

Morfessor (Creutz e Lagus, 2007): Un sistema basato su MDL che segmenta parole in morfi (unità simili a morfemi) minimizzando la somma della lunghezza del lessico morfemico e della lunghezza delle parole codificate come sequenze di morfi. Morfessor è stato applicato con successo a molte lingue e può scoprire pattern morfologici complessi senza supervisione.

Adaptor Grammars (Johnson, Griffiths, Goldwater): Estensioni Bayesiane non parametriche di grammatiche context-free che permettono di modellare dipendenze tra morfemi e di adattare le distribuzioni su strutture basate su frequenze osservate.

Questi modelli trattano implicitamente le sequenze di morfemi come processi markoviani: la probabilità di osservare una parola è data dalla probabilità della sua segmentazione morfologica, che si fattorizza come prodotto di probabilità di morfemi individuali e delle loro combinazioni.

Allomorfia e Variazione Fonologica

Un aspetto complicante della morfologia è l'allomorfia: lo stesso morfema può avere forme fonetiche diverse in contesti diversi. Per esempio, il morfema del plurale in inglese ha tre principali realizzazioni:

- /s/ dopo consonanti sorde: "cats" [kæts]
- /z/ dopo consonanti sonore o vocali: "dogs" [dɒgz], "bees" [biːz]
- /ɪz/ dopo sibilanti: "horses" [ˈhɔːsɪz]

I bambini devono apprendere sia che queste varianti rappresentano lo stesso morfema funzionale sia le regole fonotatiche che determinano quale variante occorre in quale contesto.

Un HMM fornisce un modello naturale per questo. Gli stati nascosti rappresentano morfemi astratti, e la matrice di emissione specifica come questi sono realizzati fonologicamente in diversi contesti:

$B[\text{morfema_plurale}, /s/, \text{contesto_dopo_consonante_sorda}] = \text{alta}$
 $B[\text{morfema_plurale}, /z/, \text{contesto_dopo_consonante_sonora}] = \text{alta}$
 $B[\text{morfema_plurale}, /ɪz/, \text{contesto_dopo_sibilante}] = \text{alta}$

L'apprendimento consiste nell'inferire queste mappature e nel raggruppare varianti fonologiche come realizzazioni dello stesso morfema sottostante.

Albright e Hayes (2003) hanno sviluppato un modello computazionale dettagliato di come i bambini potrebbero apprendere regole di allomorfia. Il loro "Minimal Generalization Learner" estrae pattern dalle forme osservate e genera regole che specificano come forme sottostanti sono mappate su forme superficiali, insieme a stime della loro affidabilità basate su frequenza e consistenza nell'input.

Regolarizzazione e Sovrageneralizzazione: Il Caso del Plurale

Uno dei fenomeni più studiati nell'acquisizione morfologica è la **sovragegeneralizzazione** (overregularization). I bambini inizialmente usano correttamente sia forme regolari sia forme irregolari di un paradigma inflessionali (per esempio, "went" come passato di "go" in inglese). Poi, attraversano una fase dove sovragegeneralizzano la regola regolare anche a forme irregolari, producendo errori come "goed" invece di "went". Infine, imparano a usare correttamente sia regolari sia irregolari.

Questa curva di sviluppo a forma di U è stata spiegata in modi diversi. L'approccio simbolico tradizionale postula due sistemi: un sistema di regole per forme regolari e un lessico memorizzato per irregolari. La sovrageralizzazione avviene quando il sistema di regole diventa sufficientemente forte da sovrascrivere occasionalmente forme memorizzate.

Un approccio alternativo basato su reti neurali e apprendimento statistico suggerisce un sistema unico che apprende sia regolari sia irregolari come pattern statistici. Rumelhart e McClelland (1986) hanno mostrato che una rete neurale addestrata su forme verbali inglesi esibisce spontaneamente il pattern di sovrageralizzazione a forma di U, senza distinzione esplicita tra regole e memorizzazione.

Dal punto di vista markoviano, possiamo modellare questo processo così:

Fase iniziale: L'insieme di training è piccolo e include sia regolari sia irregolari con frequenza simile. Il modello memorizza esempi specifici.

Fase intermedia: L'insieme di training si espande e la proporzione di regolari aumenta (riflettendo che nella lingua reale i verbi regolari sono molto più numerosi). Il modello estrae il pattern maggioritario (la regola regolare) e lo applica ampiamente, includendo sovrageralizzazioni.

Fase finale: Con più esperienza, il modello raffina le sue predizioni, apprendendo che certi item richiedono trattamento speciale (irregolari ad alta frequenza resistono alla regolarizzazione).

Formalmente, in un modello markoviano di formazione di parole, la probabilità di generare una forma flessa dipende dalla sua frequenza relativa nell'input e dalla sua coerenza con pattern estratti:

$$P(\text{forma_flessa} \mid \text{radice}) = \alpha \times P_{\text{memorized}}(\text{forma} \mid \text{radice}) + (1-\alpha) \times P_{\text{rule}}(\text{forma} \mid \text{radice})$$

dove α varia con esperienza e frequenza. Per item ad alta frequenza, α è alto (preferenza per memorizzazione). Per item a bassa frequenza, α è basso (preferenza per generalizzazione).

Marcus (1995) e altri hanno argomentato che modelli puramente statistici non possono catturare la natura algebrica e simbolica delle regole morfologiche. Tuttavia, lavori successivi hanno mostrato che modelli Bayesiani gerarchici che combinano apprendimento di pattern specifici e di generalizzazioni astratte possono catturare sia aspetti simbolici sia aspetti statistici.

Produttività Morfologica e Generalizzazione

Una proprietà cruciale della competenza morfologica è la **produttività**: la capacità di applicare pattern a item nuovi mai incontrati. Se imparo che il plurale di "gatto" è "gatti", posso immediatamente formare il plurale di un nuovo sostantivo come "robotino" → "robotini", anche se non ho mai sentito questa parola prima.

Gli esperimenti classici di "wug test" di Jean Berko (1958) dimostrarono che bambini di 4-5 anni possiedono questa capacità produttiva. Quando presentati con un disegno di una

creatura inventata e la descrizione "Questo è un wug", e poi mostrati due creature con "Ora ci sono due ____", i bambini producono spontaneamente "wugs", applicando correttamente la regola del plurale inglese.

Dal punto di vista dei modelli markoviani, la produttività richiede generalizzazione oltre gli esempi osservati. Questo è precisamente ciò che il smoothing e i modelli Bayesiani forniscono: invece di assegnare probabilità zero a sequenze non viste, assegnano probabilità positive basate su similarità con pattern osservati.

Un modello n-gram con smoothing può generalizzare:

$$P(\text{morfema_nuovo}) = \lambda \times P_{\text{empirica}}(\text{morfema}) + (1-\lambda) \times P_{\text{prior}}(\text{morfema})$$

dove il prior favorisce morfemi simili a quelli già noti (per esempio, stesso pattern fonologico o categoria sintattica).

Modelli Bayesiani gerarchici sono particolarmente potenti per produttività. Usando prior strutturati che favoriscono pattern regolari, il modello può estrapolare pattern da esempi limitati e applicarli produttivamente:

$$P(\text{nuova_forma} \mid \text{pattern_estratti}) \propto \sum_{\text{pattern}} P(\text{pattern} \mid \text{data}) \times P(\text{nuova_forma} \mid \text{pattern})$$

Morfologia Derivazionale e Formazione di Parole

Oltre alla morfologia inflazionale (che esprime categorie grammaticali come tempo, numero, caso), le lingue hanno morfologia derivazionale che crea nuove parole cambiando significato o categoria. Per esempio:

- "felice" (aggettivo) → "felicità" (nome) tramite suffisso -ità
- "leggere" (verbo) → "leggibile" (aggettivo) tramite suffisso -ibile
- "fare" (verbo) → "rifare" (verbo) tramite prefisso ri-

La morfologia derivazionale è tipicamente meno regolare e meno produttiva della morfologia inflazionale. Alcuni affissi derivazionali sono molto produttivi (come -ista per formare nomi di professioni: "giornalista", "violinista"), altri sono fossilizzati e non applicabili a nuove radici.

I bambini acquisiscono morfologia derivazionale più tardivamente rispetto a inflazionale, tipicamente dopo i 4-5 anni, e il processo continua attraverso l'infanzia e adolescenza.

Un modello markoviano per morfologia derivazionale deve catturare:

1. **Sequenziamento di morfemi:** Ordine di affissazione (suffissi si aggiungono dopo radice, prefissi prima)
2. **Restrizioni di selezione:** Certi affissi si attaccano solo a certe categorie di radici
3. **Produttività graduata:** Alcuni pattern sono più estendibili di altri

Formalmente:

$$P(\text{parola_derivata}) = P(\text{prefissi}) \times P(\text{radice}) \times P(\text{suffissi} \mid \text{radice}) \times P(\text{desinenze})$$

dove ogni componente è modellato con le sue distribuzioni e vincoli.

Hay e Baayen (2005) hanno studiato produttività di affissi derivazionali inglesi usando metriche basate su type/token frequency e hanno trovato che produttività è predetta da pattern statistici nell'input: affissi che appaiono con molte radici diverse (alta diversità combinatoria) sono più produttivi.

Apprendimento di Classi Inflessionali

Molte lingue hanno molteplici classi inflessionali: sottoinsiemi di parole che flettono secondo pattern diversi. Per esempio, i verbi italiani sono tradizionalmente divisi in tre coniugazioni (-are, -ere, -ire), e i sostantivi latini hanno cinque declinazioni. All'interno di ciascuna classe, l'inflessione è relativamente regolare, ma tra classi ci sono differenze.

I bambini devono apprendere:

1. Che esistono classi multiple
2. Come identificare la classe di una nuova parola
3. Come flettere parole di ciascuna classe

Questo può essere modellato come apprendimento di una **mixture of Markov models** dove ogni classe inflessionale è un modello markoviano separato, e c'è una distribuzione su classi:

$$P(\text{forma_flessa} \mid \text{radice}) = \sum_{\text{classe}} P(\text{classe} \mid \text{radice}) \times P(\text{forma} \mid \text{radice, classe})$$

L'appartenenza alla classe può dipendere da proprietà della radice (fonologia, categoria semantica).

Studi sperimentali hanno mostrato che bambini usano indizi multipli per inferire classe inflessionale di nuove parole:

- **Indizi fonologici:** Terminazioni di radici spesso predicono classe (verbi italiani in -are appartengono alla prima coniugazione)
- **Indizi semantici:** Tipo di referente può correlarsi con genere grammaticale
- **Indizi sintattici:** Contesto sintattico può disambiguare categoria

Modelli computazionali che integrano questi indizi multipli in un framework Bayesiano hanno mostrato buone performance nell'apprendimento di classi inflessionali da corpus.

Tipologia Morfologica e Variazione Cross-linguistica

Le lingue variano enormemente nella quantità e tipo di morfologia. Questa variazione pone domande su universalità vs specificità nei meccanismi di acquisizione.

Alcuni aspetti sono probabilmente universali:

- Sensibilità a regolarità statistiche in sequenze
- Capacità di segmentare unità sublessicali
- Bias per sistematicità e produttività

Altri aspetti sono specifici:

- Quali elementi sono morfemicamente decomponibili (varia da lingua a lingua)
- Ordine di morfemi (prefissazione vs suffissazione)
- Complessità del paradigma inflazionale

Studi cross-linguistici hanno trovato che bambini in lingue con morfologia ricca (turco, finlandese, quechua) acquisiscono morfologia precocemente e accuratamente, spesso producendo forme corrette prima di bambini in lingue morfologicamente povere come inglese. Questo suggerisce che salienza e regolarità della morfologia nell'input facilitano apprendimento.

Dal punto di vista markoviano, lingue con morfologia agglutinante (dove morfemi si concatenano trasparentemente) potrebbero essere più facilmente analizzabili perché confini morfemici corrispondono a cali chiari nelle probabilità di transizione. Lingue fusionali (dove morfemi si fondono) possono richiedere rappresentazioni e meccanismi di apprendimento più astratti.

Morfologia e Sintassi: L'Interfaccia

La morfologia inflazionale spesso codifica informazioni rilevanti per sintassi: caso, accordo, tempo. I bambini devono imparare non solo forme morfologiche ma anche come queste si integrano in strutture sintattiche.

Per esempio, in italiano, l'accordo di genere e numero tra articolo, aggettivo e nome:

"il gatto nero" vs "i gatti neri" vs "la gatta nera" vs "le gatte nere"

Questo richiede coordinazione tra conoscenza morfologica e sintattica. Un modello integrato rappresenta sia dipendenze sequenziali locali (morfemi all'interno di parole) sia dipendenze strutturali non-locali (accordo tra parole).

Formalmente, possiamo usare un **factorial HMM** dove esistono molteplici catene markoviane parallele:

- Una catena per sequenza di radici lessicali
- Catene separate per features morfosintattiche (numero, genere, caso)
- Vincoli di coerenza tra catene (accordo)

Questo framework permette di modellare interazioni complesse tra morfologia e sintassi mantenendo trattabilità computazionale.

Morfologia Processuale e Computazionale

Oltre a modellare acquisizione, i modelli markoviani sono rilevanti per comprendere l'elaborazione morfologica online. Quando comprendiamo o produciamo parole morfologicamente complesse, le decomponiamo in morfemi o le trattiamo come unità olistiche?

Evidenze sperimentali (priming morfologico, potenziali evento-correlati, eye-tracking) suggeriscono che parole morfologicamente complesse sono decomposte durante elaborazione, almeno per forme trasparenti e regolari. Per forme frequenti o irregolari, può esserci elaborazione olistica.

Un modello markoviano di elaborazione morfologica potrebbe rappresentare questo come competizione tra due vie:

1. **Via decompositiva:** Segmenta parola in morfemi, recupera significati di morfemi, compone significato
2. **Via olistica:** Recupera parola intera dal lessico mentale

La via dominante dipende da fattori come frequenza della parola intera, trasparenza semantica della composizione, regolarità della formazione.

Questo è analogo ai modelli dual-route per lettura di parole, e può essere implementato come combinazione di modelli markoviani per decomposizione e modelli di recupero diretto.

Implicazioni per Educazione e Patologia

Comprendere acquisizione morfologica attraverso modelli computazionali ha implicazioni pratiche:

Per insegnamento di lingue: Sapere che apprendimento morfologico è facilitato da input regolare e frequente suggerisce strategie pedagogiche che enfatizzano pattern ricorrenti e pratica produttiva.

Per valutazione di disturbi: Bambini con SLI spesso mostrano deficit specifici in morfologia, particolarmente infrazioni regolari. Modelli computazionali possono aiutare a diagnosi quantificando gravità dei deficit e identificando pattern specifici di difficoltà.

Per interventi terapeutici: Approcci basati su evidenze che enfatizzano esposizione massiva a esemplari e feedback su generalizzazioni sono coerenti con principi di apprendimento statistico.

La morfologia, vista attraverso lente dei processi markoviani, emerge come dominio dove regolarità statistiche nell'input interagiscono con bias cognitivi per produttività e sistematicità, risultando in sistemi che sono allo stesso tempo pattern-based e algebrici, memorizzati e computati, specifici agli item e generalizzabili, riflettendo la complessità straordinaria del linguaggio umano e della cognizione che lo supporta.

Parte V: Dai Pattern Sequenziali alla Sintassi: Transizioni Categoriali e Primi Passi Grammaticali

Il Problema della Sintassi: Oltre le Sequenze Lineari

Mentre i modelli markoviani hanno dimostrato notevole successo nella modellazione dell'acquisizione fonologica e morfologica, l'applicazione alla sintassi è più controversa. La sintassi delle lingue naturali esibisce proprietà che sembrano andare oltre semplici

dipendenze sequenziali: ricorsività illimitata, dipendenze a lunga distanza, isole sintattiche, e struttura gerarchica che non è riducibile a ordine lineare.

Chomsky (1957) ha famoso argumentato che le lingue naturali non possono essere generate da grammatiche a stati finiti (equivalenti a modelli markoviani). Costruzioni come dipendenze annidate centrate (center-embedded dependencies) richiedono memoria strutturata che va oltre le capacità di catene di Markov finite:

"Il topo che il gatto che il cane inseguiva cacciava mangiò il formaggio"

Per processare correttamente questa frase, dobbiamo tenere traccia di molteplici dipendenze aperte simultaneamente in modo strutturalmente organizzato, non meramente come sequenza di stati.

Tuttavia, questo non significa che i modelli markoviani siano irrilevanti per l'acquisizione sintattica. Piuttosto, suggerisce che rappresentano un componente di un sistema più complesso. In particolare:

1. Nelle fasi iniziali dell'acquisizione, i bambini potrebbero utilizzare strategie basate su pattern sequenziali come primo passo
2. Anche la sintassi adulta contiene regolarità statistiche sequenziali che i modelli markoviani possono catturare
3. Pattern markoviani potrebbero servire come "impalcatura" (scaffold) su cui si costruiscono rappresentazioni più astratte

Proto-Sintassi: Le Prime Combinazioni di Parole

I bambini iniziano tipicamente a produrre singole parole isolate intorno ai 12 mesi. Tra 18 e 24 mesi, cominciano a combinare due parole in enunciati come "mamma pappa", "bimbo caduto", "ancora biscotto". Questi primi enunciati biparola mostrano pattern interessanti che sono stati estensivamente studiati.

Braine (1963) ha descritto le "parole pivot" (pivot words): un piccolo insieme di parole che appaiono frequentemente in posizione fissa (tipicamente prima o dopo) in combinazione con un insieme più ampio di parole aperte. Per esempio:

Pivot iniziali: "più x", "altro x", "no x" (dove x è un nome qualsiasi) Pivot finali: "x via", "x qui", "x giù"

Questo pattern può essere facilmente modellato come una catena di Markov di primo ordine con due stati: {Pivot, Aperto}. Le transizioni permesse sono limitate:

$P(\text{Aperto} \mid \text{Pivot}) = \text{alta}$ $P(\text{Fine} \mid \text{Aperto}) = \text{alta}$ $P(\text{Pivot} \mid \text{Aperto}) = \text{bassa}$ (le pivot words non si seguono tipicamente)

Critiche successive hanno mostrato che la grammatica pivot non cattura completamente la complessità anche degli enunciati biparola (esistono combinazioni Aperto-Aperto, relazioni semantiche diverse), ma l'osservazione di base rimane: esistono pattern sequenziali regolari nelle prime produzioni.

Modelli N-gram e Sintassi Iniziale

Un modello bigram semplice per proto-sintassi assegna probabilità a sequenze di due parole:

$$P(w_1, w_2) = P(w_1) \times P(w_2 | w_1)$$

Se addestrato su produzioni infantili o su parlato infantile-diretto, questo modello cattura molte delle combinazioni frequenti e può persino "generare" enunciati bipolarola plausibili campionando dalla distribuzione appresa.

Reali, Christiansen e Monaghan (2003) hanno addestrato modelli n-gram su corpus di parlato infantile-diretto e hanno trovato che modelli trigram (ordine 2) potevano predire categorie grammaticali di parole nelle frasi con accuratezza sorprendentemente alta (circa 80-90%), basandosi esclusivamente su informazioni di co-occorrenza. Questo suggerisce che informazioni distribuzionali codificate in modelli markoviani catturano molto sulla struttura categoriale della lingua.

Bootstrapping Distribuzionale: Dalle Sequenze alle Categorie

Un meccanismo cruciale proposto per acquisizione sintattica è il **bootstrapping distribuzionale** (distributional bootstrapping): i bambini usano informazioni su come le parole si distribuiscono in contesti linguistici per inferire le loro categorie grammaticali.

L'idea risale a Harris (1951) e Firth (1957): parole che appaiono in contesti simili tendono a essere simili sintatticamente e semanticamente. Per esempio, se osserviamo gli schemi:

"il ____ corre" "un ____ salta" "questo ____ mangia"

possiamo inferire che qualsiasi parola che può riempire lo spazio vuoto è probabilmente un nome (o più precisamente, una testa nominale).

Formalmente, possiamo rappresentare ciascuna parola come un vettore di contesti in cui appare. Due parole sono distribuzionalmente simili se i loro vettori di contesto sono simili (alta correlazione o bassa distanza).

Un modello markoviano cattura questo attraverso probabilità condizionali:

$$\text{Similarità}(w_1, w_2) \propto \text{correlazione}(P(\text{context} | w_1), P(\text{context} | w_2))$$

Redington, Chater e Finch (1998) hanno dimostrato che clustering basato su similarità distribuzionale di parole in corpus di parlato infantile-diretto produce clusters che corrispondono approssimativamente a categorie grammaticali tradizionali (nomi, verbi, aggettivi), senza alcuna supervisione esplicita.

Mintz (2003) ha introdotto i **frequent frames**: sequenze non continue come "il ____ è" o "tu ____ lo". Parole che appaiono nello stesso frame frequente tendono fortemente a condividere categoria grammaticale. Bambini di 12 mesi sono sensibili a frequent frames e li usano per categorizzare parole nuove in esperimenti di laboratorio, suggerendo che questo è un meccanismo cognitivamente reale.

Hidden Markov Models per Inducing Grammatiche

Gli HMM forniscono un framework naturale per modellare l'induzione di categorie sintattiche (parti del discorso) da testo non annotato. In questa applicazione:

- **Stati nascosti** = categorie grammaticali (nome, verbo, aggettivo, ecc.)
- **Osservazioni** = parole concrete
- **Matrice di transizione** $A[i,j]$ = probabilità che categoria j segua categoria i
- **Matrice di emissione** $B[i,w]$ = probabilità che categoria i emetta parola w

Dato un corpus non annotato, l'algoritmo di Baum-Welch può essere usato per apprendere parametri (A e B) e per inferire la sequenza più probabile di categorie nascoste per ogni frase (usando Viterbi).

Merialdo (1994) e successivi ricercatori hanno mostrato che HMM addestrati su grandi corpora possono indurre categorie grammaticali che corrispondono a Part-Of-Speech tags con accuratezza ragionevole (70-80% senza supervisione, oltre 95% con supervisione parziale).

Tuttavia, le categorie indotte non sono sempre interpretabili come categorie grammaticali tradizionali. Spesso riflettono distinzioni più fini (nomi comuni vs nomi propri vs pronomi) o più grossolane (tutte le parole chiuse insieme). Questo solleva domande su se le categorie indotte computazionalmente corrispondono a quelle rappresentate cognitivamente dai bambini.

Limitazioni dei Modelli Markoviani per Sintassi Complessa

Nonostante successi nell'induzione di categorie e nella cattura di regolarità sequenziali locali, i modelli markoviani di ordine fisso hanno limitazioni ben note per sintassi complessa:

Dipendenze a lunga distanza: In "Il libro che Maria ha letto era interessante", "libro" è soggetto di "era" ma sono separati da una frase relativa. Un bigram o trigram non può catturare questa dipendenza diretta.

Ricorsività: Le lingue permettono incassamento arbitrario di clausole, che richiede memoria potenzialmente illimitata. Catene di Markov con numero finito di stati non possono generare linguaggi context-free generali.

Dipendenze non-locali e movimento: Fenomeni come movimento wh-, topicalizzazione, anaforica a lunga distanza richiedono rappresentazioni strutturali che vanno oltre semplici stati sequenziali.

Isole sintattiche: Certe dipendenze sono impossibili anche se sarebbero sequenzialmente plausibili, indicando vincoli strutturali gerarchici non catturabili in modelli puramente sequenziali.

Queste limitazioni hanno portato allo sviluppo di modelli più sofisticati.

Estensioni: Modelli Markoviani Strutturati

Diverse estensioni dei modelli markoviani sono state proposte per catturare aspetti strutturali della sintassi mantenendo trattabilità computazionale.

Grammatiche Probabilistiche Context-Free (PCFG): Generalizzano modelli markoviani permettendo generazione gerarchica di strutture ad albero. Una PCFG specifica probabilità per regole di riscrittura:

$A \rightarrow \alpha$ [probabilità p]

Le PCFG possono modellare ricorsività e struttura costituente, mantenendo framework probabilistico. Tuttavia, pure PCFG hanno difficoltà con dipendenze lessicali e fenomeni non-context-free.

Tree-Adjoining Grammars (TAG): Sistemi più potenti che permettono incassamento e adjunction di strutture arboree, catturando meglio fenomeni come movimento e dipendenze a lunga distanza.

Data-Oriented Parsing (DOP): Approccio dove probabilità sono assegnate non a regole atomiche ma a frammenti arbitrari di alberi sintattici osservati nel corpus. Questo cattura idiosincrasie e pattern multi-livello.

Modelli di Dipendenza Probabilistici: Come discusso nel saggio precedente, modellano direttamente dipendenze tra parole piuttosto che struttura costituente. Possono essere visti come generalizzazione di modelli markoviani dove stati corrispondono a configurazioni di dipendenze parziali.

Apprendimento Sintattico Bayesiano

Un approccio influente è l'uso di modelli Bayesiani gerarchici per acquisizione sintattica. Questi modelli specificano prior su grammatiche possibili e usano inferenza Bayesiana per aggiornare credenze basate sull'input osservato.

Perfors, Tenenbaum e Regier (2011) hanno esplorato acquisizione di grammatiche context-free usando Bayesian Model Selection. Il loro modello considera uno spazio di grammatiche possibili e assegna probabilità posteriori basate su quanto bene ciascuna grammatica spiega i dati osservati e quanto è semplice (prior di parsimonia).

Risultato cruciale: il modello mostra preferenza per grammatiche gerarchiche su grammatiche basate su semplice adiacenza sequenziale, anche quando entrambe spiegano ugualmente i dati superficiali. Questo perché grammatiche gerarchiche sono intrinsecamente più semplici (meno parametri necessari per specificarle). Questo suggerisce una spiegazione computazionale per bias verso struttura gerarchica nell'acquisizione sintattica.

Ruolo dell'Input: Quantità vs Qualità

Importante domanda: quanto conta la quantità di input nell'acquisizione sintattica? Studi su variazione socioeconomica hanno trovato che bambini da famiglie con status socioeconomico più alto sono esposti a significativamente più parlato, e questo predice differenze nel vocabolario e nella grammatica.

Dal punto di vista di modelli statistici, più input significa stima più accurata di probabilità e più opportunità per estrarre pattern rari. Simulazioni hanno mostrato che modelli addestrati su quantità diverse di input mostrano curve di apprendimento che qualitativamente corrispondono a sviluppo osservato.

Tuttavia, qualità dell'input è anche cruciale. Parlato infantile-diretto (child-directed speech, CDS) ha proprietà specifiche: pitch più alto, esagerazione prosodica, maggiore ripetitività, frasi più brevi e grammaticalmente più semplici. Queste proprietà potrebbero facilitare apprendimento statistico rendendo pattern più salienti.

Esperimenti hanno mostrato che CDS effettivamente facilita apprendimento di pattern grammaticali in laboratorio. Modelli computazionali addestrati su CDS apprendono più rapidamente rispetto a modelli addestrati su parlato adulto-diretto.

Interazione tra Apprendimento Statistico e Conoscenze Innate

Dibattito centrale: quanto dell'acquisizione sintattica può essere spiegato attraverso apprendimento statistico dominio-generale, e quanto richiede conoscenze o bias linguistici specifici?

Posizioni estreme:

Innatismo forte: I bambini nascono con conoscenza sostanziale di principi grammaticali universali. Apprendimento consiste principalmente nell'impostare parametri binari (per esempio, ordine testa-complemento).

Empirismo estremo: I bambini nascono solo con meccanismi di apprendimento statistico generali. Tutta la conoscenza grammaticale è estratta dall'input attraverso tracciamento di regolarità.

Posizioni intermedie (probabilmente più plausibili):

Bias induttivi leggeri: I bambini hanno predisposizioni per certi tipi di rappresentazioni (gerarchiche, categoriali) e certi bias di apprendimento (preferenza per semplicità, coerenza), ma non conoscenze grammaticali specifiche.

Meccanismi dominio-general con specializzazione: Meccanismi computazionali generali (tracciamento statistico, chunking, inferenza strutturale) si specializzano attraverso apprendimento per elaborazione linguistica.

Evidenze recenti da neuroimaging suggeriscono che le reti cerebrali per elaborazione linguistica si sovrappongono parzialmente ma non completamente con reti per altri domini cognitivi, supportando specializzazione acquisita piuttosto che moduli completamente innati.

Modelli Neurocomputazionali: Collegare Markov e Meccanismi Neurali

Come potrebbero catene di Markov o HMM essere implementate neuralmente? Diversi meccanismi neurali sono candidati:

Assemblies di neuroni con plasticità hebbiana: Neuroni che si attivano insieme ripetutamente sviluppano connessioni forti. Sequenze frequenti nel linguaggio creano "catene" di assemblies con connessioni facilitate, implementando implicitamente probabilità di transizione.

Oscillazioni neurali e binding temporale: Diverse fasi di oscillazioni potrebbero codificare diversi stati in una sequenza, con sincronizzazione tra regioni che implementa transizioni di stato.

Reti ricorrenti e memoria: Reti neurali ricorrenti (RNN) e loro varianti come LSTM possono apprendere dipendenze temporali complesse e possono essere viste come generalizzazioni di modelli markoviani con memoria strutturata.

Codifica predittiva: Il cervello come "macchina di predizione" che continuamente genera aspettative su input futuro basate su input passato. Errori di predizione guidano apprendimento. Questo è naturalmente modellato come processo markoviano dove predizioni sono probabilità di transizione apprese.

Modelli computazionali biologicamente plausibili che implementano questi meccanismi sono stati sviluppati e mostrano capacità di apprendere regolarità sintattiche da input simulato, suggerendo vie per connettere livelli computazionale e implementazionale.

Sviluppo Tipico e Atipico: Variazioni nei Meccanismi Markoviani

Studiare popolazioni con sviluppo atipico fornisce insights su componenti del sistema di acquisizione.

Disturbo Specifico del Linguaggio (SLI): Bambini con SLI mostrano particolare difficoltà con morfosintassi ma intelligenza non-verbale preservata. Alcuni ricercatori hanno proposto che SLI riflette deficit in apprendimento di pattern sequenziali rapidi. Esperimenti con compiti di apprendimento statistico implicito (non linguistici) hanno trovato che bambini con SLI performano più debolmente, supportando ipotesi di deficit in meccanismi markoviani generali.

Disturbi dello Spettro Autistico (ASD): Profili linguistici in ASD sono eterogenei. Alcuni individui con ASD mostrano apprendimento lessicale forte ma difficoltà con pragmatica e aspetti sociali del linguaggio. Apprendimento statistico di pattern sequenziali può essere preservato o persino superiore, ma integrazione con informazione sociale e contestuale è compromessa.

Sindrome di Williams: Profilo opposto: linguaggio relativamente preservato (fluente, grammaticale) ma cognizione non-linguistica compromessa. Questo è stato interpretato come evidenza per modularità del linguaggio, sebbene analisi dettagliate mostrino che anche il linguaggio ha sottili deficit, particolarmente in aspetti complessi.

Modellare queste popolazioni richiede variare parametri nei modelli computazionali (per esempio, tasso di apprendimento, capacità di memoria, bias attenzionali) per vedere quali patterns di deficit emergono, fornendo ipotesi testabili su meccanismi sottostanti.

Transizione da Statistiche Implicite a Conoscenze Esplicite

Un aspetto affascinante dell'acquisizione sintattica è che conoscenza che è inizialmente implicita (operazionale, procedurale) diventa gradualmente più esplicita e accessibile a introspezione e manipolazione conscia, particolarmente con educazione formale.

Bambini piccoli usano grammatica correttamente molto prima di poter articolare regole grammaticali. Giudizi di grammaticalità diventano sempre più accurati e sofisticati attraverso sviluppo. Capacità metalinguistiche (riflettere su linguaggio come oggetto) emergono gradualmente.

Questa transizione può essere modellata come passaggio da rappresentazioni distribuite e probabilistiche (codificate in modelli markoviani neuralmente implementati) a rappresentazioni più simboliche e strutturate (accessibili a elaborazione cognitiva di livello superiore). Meccanismi di consolidamento e riorganizzazione della conoscenza durante sonno e sviluppo potrebbero giocare ruoli cruciali.

Sintassi come Sistema Multi-Livello

Sintesi emergente: la sintassi delle lingue naturali non è né puramente sequenziale (come catturerebbe un modello markoviano semplice) né puramente simbolica/gerarchica (come postulerebbe teoria generativa classica), ma piuttosto un sistema multi-livello dove:

- **Livello inferiore:** Regolarità sequenziali, co-occorrenze, chunking statistico (bene modellato da approcci markoviani)
- **Livello intermedio:** Categorie grammaticali, costruzioni, schemi (modellabili con HMM, PCFG, Construction Grammar)
- **Livello superiore:** Principi astratti, dipendenze strutturali, ricorsività (richiedente rappresentazioni gerarchiche esplicite)

Apprendimento procede dal basso verso l'alto: inizia con sensibilità a pattern sequenziali, costruisce rappresentazioni categoriali intermedie attraverso bootstrapping distribuzionale, e infine astrae principi strutturali generali. Questo è un processo graduale e incrementale, non un salto discontinuo.

Modelli computazionali contemporanei (particolarmente quelli basati su reti neurali profonde come Transformers) catturano aspetti di questa architettura multi-livello, apprendendo simultaneamente rappresentazioni a diversi livelli di astrazione attraverso gerarchie di elaborazione. Questi modelli mostrano capacità sintattiche impressionanti pur essendo addestrati principalmente su obiettivi predittivi sequenziali, suggerendo che rappresentazioni strutturali possono emergere da apprendimento su regolarità statistiche quando architetture computazionali appropriate sono impiegate.

I modelli markoviani, lungi dall'essere obsoleti per sintassi, rappresentano una base essenziale su cui si costruiscono competenze sintattiche più complesse. Forniscono meccanismi per tracciare pattern, formare categorie, e estrarre regolarità che sono precursori e componenti di rappresentazioni grammaticali mature. Comprenderli è quindi cruciale per una teoria completa dell'acquisizione sintattica.

Parte VI: Prospettive Teoriche, Implementazione Neurale e Sintesi Conclusiva

Il Dibattito tra Approcci Simbolici e Statistici

La storia della scienza cognitiva e della linguistica computazionale è stata segnata da un dibattito profondo tra due visioni contrastanti della rappresentazione e dell'elaborazione linguistica: l'approccio simbolico-computazionale e l'approccio connessionista-statistico.

L'**approccio simbolico**, dominante nella linguistica generativa e nell'intelligenza artificiale classica, postula che la competenza linguistica consiste di regole discrete che operano su rappresentazioni simboliche strutturate. Secondo questa visione, apprendere il passato dei verbi in inglese significa acquisire una regola algebrica del tipo:

Verbo + /ed/ → Passato

insieme a un lessico di eccezioni memorizzate. Le rappresentazioni sono composizionali, sistematiche, e produttive in virtù della loro struttura simbolica.

L'**approccio connessionista**, emerso con forza negli anni Ottanta, propone che la conoscenza linguistica è codificata in pesi di connessioni in reti neurali distribuite. Non esistono regole esplicite o rappresentazioni simboliche discrete, ma piuttosto pattern di attivazione attraverso unità di elaborazione. Apprendimento consiste in aggiustamenti graduali di pesi basati su esposizione a esempi.

I modelli markoviani occupano una posizione interessante in questo dibattito. Sono statistici nel senso che codificano probabilità e sono appresi da dati, ma mantengono un certo grado di struttura simbolica (stati discreti, transizioni definite). Possono essere visti come punto intermedio o come possibile riconciliazione tra i due approcci.

Argomenti a Favore di Approcci Statistici nell'Acquisizione

Diversi argomenti supportano il ruolo centrale di meccanismi statistici come modelli markoviani nell'acquisizione:

Evidenze sperimentali: Come ampiamente documentato, bambini dimostrano sensibilità precoce e robusta a pattern statistici in stimoli linguistici e non-linguistici. Questa capacità è osservabile nei primi mesi di vita, ben prima di competenza linguistica esplicita.

Plausibilità neurologica: Meccanismi di plasticità sinaptica che implementerebbero apprendimento statistico (potenziamento a lungo termine, STDP) sono ben documentati in neuroscienze. Nessun analogo neurobiologico convincente è stato proposto per manipolazione di simboli discreti in stile computazione classica.

Parsimonia evolutiva: Meccanismi di apprendimento statistico dominio-generalmente potrebbero essere evolutivamente più antichi e condivisi con altre specie. Postulare meccanismi specificamente linguistici richiede spiegazione della loro origine evolutiva.

Successo computazionale: Modelli statistici contemporanei (inclusi modelli basati su deep learning) raggiungono performance impressionanti su compiti linguistici complessi, spesso superando sistemi basati su regole esplicite.

Gradualità e continuità: Sviluppo linguistico infantile mostra gradualità e continuità piuttosto che salti discontinui che ci si aspetterebbe se emergesse da maturazione di moduli innati.

Argomenti per Complementarietà e Interazione

Tuttavia, argomenti persuasivi suggeriscono che apprendimento statistico da solo è insufficiente e deve essere complementato da altri meccanismi:

Poverty of stimulus: Sebbene attenuato da analisi computazionali moderne, rimane il fatto che bambini convergono su generalizzazioni che sembrano sottodeterminate dall'input. Per esempio, bambini rispettano vincoli su strutture interrogative (isole sintattiche) senza evidenza diretta negativa nell'input.

Bias universali: Esistono pattern sintattici che sono virtualmente universali nelle lingue del mondo (per esempio, preferenze per ordini basici di soggetto, oggetto, verbo; restrizioni su movimenti sintattici). Questi potrebbero riflettere bias innati piuttosto che convergenza indipendente.

Limiti della generalizzazione statistica: Puro tracciamento di frequenze non spiega facilmente generalizzazioni astratte che vanno oltre similarità superficiale. Per esempio, bambini applicano regole morfologiche a non-parole con struttura fonologica mai incontrata.

Specificità del linguaggio: Sebbene apprendimento statistico sia dominio-generale, esistono anche specializzazioni neurali e comportamentali per linguaggio che emergono precocemente e che suggeriscono qualcosa di più di semplice applicazione di meccanismi generali.

La sintesi più promettente sembra essere che acquisizione linguistica coinvolge interazione stretta tra meccanismi statistici dominio-general (ben modellati da approcci markoviani) e bias o vincoli specifici (possibilmente innati o acquisiti precocemente attraverso interazione con architettura percettivo-motoria).

Modelli Bayesiani Come Framework Unificante

I modelli Bayesiani hanno emergentemente come framework potenzialmente unificante che integra apprendimento da dati con conoscenze o bias a priori. In formulazione Bayesiana:

$$P(\text{Ipotesi} \mid \text{Dati}) \propto P(\text{Dati} \mid \text{Ipotesi}) \times P(\text{Ipotesi})$$

Il **posterior** (credenza aggiornata) è determinato sia dal **likelihood** (quanto i dati sono probabili data l'ipotesi, componente empirica/statistica) sia dal **prior** (quanto l'ipotesi era plausibile a priori, componente innata/bias).

Questo framework permette di formulare dibattiti classici in termini quantitativi:

- **Innatismo vs empirismo:** Non più questione binaria ma questione di forza relativa dei prior vs evidenza empirica
- **Universali linguistici:** Possono emergere da prior comuni o da convergenza statistica su soluzioni ottimali

- **Differenze individuali:** Variazioni in prior, in evidenza ricevuta, o in processi di inferenza

Studi empirici hanno iniziato a testare predizioni di modelli Bayesiani. Per esempio, se i bambini fanno inferenza Bayesiana, dovrebbero essere più disposti ad adottare generalizzazioni che sono a priori plausibili anche con evidenza limitata, ma dovrebbero resistere a generalizzazioni implausibili anche con evidenza moderata. Esperimenti mostrano pattern coerenti con questo.

Implementazione Neurale di Processi Markoviani

Come sono implementati processi markoviani nel cervello? Diverse linee di evidenza informano questa questione.

Codifica di sequenze temporali: Studi di neurofisiologia hanno identificato neuroni che rispondono selettivamente a specifiche sequenze temporali di stimoli. Questi "neuroni di sequenza" sono stati trovati in ippocampo, corteccia prefrontale, e altre aree coinvolte in memoria e pianificazione.

Plasticità dipendente dal tempo: STDP (Spike-Timing Dependent Plasticity) fornisce un meccanismo cellulare attraverso cui connessioni sinaptiche si rafforzano quando neuroni pre-sinaptici si attivano ripetutamente poco prima di neuroni post-sinaptici. Questo crea naturalmente catene di attivazione che codificano transizioni frequenti, implementando implicitamente modello markoviano.

Oscillazioni e dinamica neurale: Pattern oscillatori in diverse bande di frequenza (theta, gamma) coordinano temporalmente attività neurale. Si è proposto che diverse fasi di oscillazioni potrebbero separare rappresentazioni di elementi sequenziali successivi, con transizioni di fase che implementano transizioni di stato.

Attractor networks: Reti neurali ricorrenti possono formare "attrattori" dove certi pattern di attività sono stabili. Transizioni tra attrattori implementano naturalmente catene di Markov dove stati discreti sono pattern di attività stabili.

Codifica predittiva: Framework influente propone che il cervello implementa inferenza Bayesiana gerarchica attraverso comparazione continua tra predizioni top-down e input bottom-up. Errori di predizione guidano sia percezione sia apprendimento. In questo framework, apprendimento di regolarità statistiche (come transizioni markoviane) emerge naturalmente.

Simulazioni neurocomputazionali che implementano questi meccanismi hanno dimostrato capacità di apprendere regolarità statistiche simili a quelle osservate in studi comportamentali di apprendimento statistico infantile, fornendo prova di principio per plausibilità biologica.

Variazioni Cross-linguistiche e Universali

Una predizione di approcci fortemente statistici è che pattern di acquisizione dovrebbero riflettere strutture statistiche specifiche di ciascuna lingua. Una predizione di approcci

innatisti è che dovrebbero esistere universali nell'acquisizione indipendenti da strutture linguistiche specifiche.

Evidenze empiriche mostrano entrambi i pattern:

Variazione: Bambini in lingue diverse mostrano tempistiche diverse per acquisizione di diversi fenomeni. Per esempio, bambini in lingue con morfologia ricca e regolare (turco, finlandese) acquisiscono morfologia inflessionale prima di bambini in inglese. Questo è coerente con apprendimento guidato da statistiche dell'input.

Universali: Certe sequenze di sviluppo sono comuni attraverso lingue. Per esempio, quasi universalmente i bambini producono affermative prima di negative, attive prima di passive, frasi semplici prima di complesse. Alcuni di questi universali potrebbero riflettere non tanto bias innati quanto vincoli cognitivi generali (limitazioni di memoria di lavoro, processamento incrementale da sinistra a destra).

Slobin (1985) ha proposto "principi operativi" che guidano acquisizione: bias per processare stimoli salienti percettivamente, prestare attenzione a finali di parole, cercare sistematicità. Questi principi sono dominio-general ma hanno conseguenze specifiche per linguaggio date strutture delle lingue.

Modelli computazionali che incorporano tali bias insieme ad apprendimento statistico mostrano migliore fit a dati cross-linguistici rispetto a modelli puramente statistici o puramente parametrici.

Apprendimento Multimodale e Embodiment

Un'estensione importante di modelli markoviani standard è considerare che l'input linguistico non avviene in isolamento ma è integrato con informazioni visive, gestuali, e situazionali. I bambini non sentono solo parole ma vedono referenti, osservano azioni, e partecipano a interazioni sociali.

Apprendimento cross-modale: Le probabilità di transizione non sono solo tra elementi linguistici ma tra elementi linguistici e percetti non-linguistici. Per esempio:

$P(\text{parola} \mid \text{oggetto_visivo}, \text{contesto})$ $P(\text{azione} \mid \text{verbo}, \text{situazione})$

Modelli che incorporano queste co-occorrenze cross-modal mostrano migliori prestazioni in compiti come apprendimento di mappature parola-referente e comprensione di verbi di azione.

Grounding sensorimotorio: Le rappresentazioni linguistiche potrebbero essere "grounded" in sistemi percettivi e motori. Comprendere la parola "afferrare" potrebbe attivare aree motorie coinvolte nell'afferrare realmente. Modelli embodied di cognizione suggeriscono che significato linguistico è costituito da simulazioni sensorimotorie.

Dal punto di vista markoviano, questo significa estendere spazio degli stati per includere non solo simboli linguistici ma anche stati percettivi, motori, e attenzionali. L'apprendimento diventa un processo di tracciare co-occorrenze e transizioni attraverso questo spazio multimodale espanso.

Evidenze sperimentali mostrano che bambini sfruttano informazioni cross-modali: lo sguardo congiunto (joint attention) facilita apprendimento di parole; gesti informativi sincronizzati con parlato migliorano comprensione; manipolazione fisica di oggetti durante esposizione linguistica migliora ritenzione.

Differenze Individuali e Traiettorie Multiple

Non tutti i bambini seguono esattamente la stessa traiettoria di acquisizione. Esistono differenze individuali sostanziali nella tempistica, nel percorso, e nell'outcome finale. Comprendere fonti di questa variazione è cruciale.

Variazione nell'input: Quantità e qualità di linguaggio a cui bambini sono esposti varia enormemente. Questo predice differenze in tasso di acquisizione in modelli statistici.

Variazione in capacità cognitive: Differenze in memoria di lavoro, velocità di elaborazione, attenzione sostenuta si correlano con differenze in acquisizione linguistica.

Stili di apprendimento: Alcuni bambini sono "referenziali" (focalizzati su nomi e oggetti), altri "espressivi" (focalizzati su funzioni sociali). Questo potrebbe riflettere differenze in strategie attenzionali o motivazionali.

Fattori socio-emotivi: Qualità di interazione sociale, responsività genitoriale, e contesto emotivo influenzano apprendimento.

Modelli computazionali possono simulare queste differenze variando parametri (tasso di apprendimento, capacità di memoria, dimensione di finestra attenzionale) o variando caratteristiche dell'input. Questo permette di generare ipotesi testabili su meccanismi sottostanti a differenze individuali.

Importante insight: esistono molteplici percorsi verso competenza linguistica. Non c'è necessariamente un'unica sequenza "corretta" di acquisizione. Diversi meccanismi e strategie possono portare a outcome funzionalmente equivalenti.

Implicazioni per Intervento e Educazione

Comprendere acquisizione linguistica attraverso lente di modelli markoviani ha implicazioni pratiche.

Identificazione precoce di ritardi: Misure di sensibilità ad apprendimento statistico a età precoce (6-8 mesi) predicono sviluppo linguistico successivo. Questo potrebbe permettere identificazione di bambini a rischio prima che ritardi diventino evidenti.

Interventi mirati: Se deficit specifici in tracciamento di statistiche temporali contribuiscono a disturbi linguistici, interventi che rafforzano queste capacità potrebbero essere benefici. Training computerizzato su compiti di apprendimento sequenziale ha mostrato promesse preliminari.

Ottimizzazione dell'input: Sapere che apprendimento è guidato da statistiche dell'input suggerisce che aumentare sistematicità, salienza, e ripetizione di pattern target nell'input

facilita acquisizione. Questo è già pratica in terapia del linguaggio ma può essere raffinata basandosi su principi computazionali.

Apprendimento di seconda lingua: Adulti che apprendono seconda lingua possono beneficiare da training esplicito su pattern statistici (probabilità di transizione fonologiche, collocazioni frequenti) per accelerare formazione di intuizioni native.

Tecnologie assistive: Sistemi di predizione di testo, di riconoscimento vocale, e di sintesi vocale basati su modelli markoviani possono essere personalizzati per individui con bisogni comunicativi speciali.

Direzioni Future di Ricerca

Il campo dell'acquisizione linguistica computazionale è in rapida evoluzione. Diverse direzioni promettenti:

Modelli neurali biologicamente plausibili: Integrare più strettamente modelli computazionali con conoscenze neurobiologiche dettagliate su architetture circuitali, dinamiche oscillatorie, neuromodulazione.

Apprendimento continuo e incrementale: La maggior parte dei modelli è addestrata batch su corpus statici. Bambini reali apprendono continuamente da stream di esperienze. Modelli di lifelong learning e online learning sono più realistici.

Integrazione di apprendimento supervisionato e non-supervisionato: Bambini ricevono sia input non-supervisionato (semplice esposizione) sia feedback esplicito (correzioni, espansioni). Modelli che integrano entrambi potrebbero essere più potenti.

Modelli multi-agente e apprendimento sociale: Acquisizione linguistica avviene in contesto sociale con molteplici parlanti che interagiscono. Modelli che simulano dinamiche sociali potrebbero catturare aspetti attualmente trascurati.

Connecting levels: Collegare più strettamente livelli computazionale (algoritmi), algoritmico (rappresentazioni), e implementazionale (neurobiologia) come proposto da framework di Marr.

Cross-linguistic computational studies: Applicare sistematicamente modelli a molte lingue tipologicamente diverse per identificare universali computazionali e parametri di variazione.

Sintesi Conclusiva: Il Linguaggio Come Sistema Statistico Strutturato

L'applicazione di modelli markoviani e processi stocastici all'acquisizione del linguaggio ha rivelato che il linguaggio umano, pur essendo straordinariamente complesso e ricco, è anche profondamente strutturato da regolarità statistiche a molteplici livelli. Dal ritmo delle sillabe nel parlato continuo, alle distribuzioni di fonemi nelle parole, alle co-occorrenze di morfemi, alle transizioni tra categorie sintattiche, i pattern probabilistici permeano ogni aspetto della struttura linguistica.

I bambini, lungi dall'essere "tabule rasae" passive o esecutori di programmi innati rigidi, emergono come straordinariamente sofisticati "statistici intuitivi" che estraggono, rappresentano e sfruttano regolarità nell'input ambientale. Questa capacità di apprendimento statistico, supportata da meccanismi neurali come plasticità sinaptica dipendente dal tempo e coding predittivo, fornisce una base potente su cui si costruisce competenza linguistica.

Tuttavia, apprendimento statistico non opera in vuoto. È modulato da:

Bias percettivi e attenzionali: Non tutte le informazioni nell'input sono ugualmente accessibili. Salienza percettiva, limitazioni di memoria, e preferenze attenzionali plasmano ciò che viene appreso.

Vincoli strutturali: Preferenze per rappresentazioni gerarchiche, categoriali, e sistematiche guidano induzione di strutture grammaticali che vanno oltre semplici associazioni sequenziali.

Integrazione multimodale: Linguaggio è appreso in contesti ricchi che integrano informazioni linguistiche, percettive, motorie, e sociali. Questa integrazione amplifica e vincola apprendimento.

Sviluppo dinamico: Acquisizione non è processo statico ma sistema dinamico dove conoscenze emergenti a un livello facilitano e vincolano apprendimento ad altri livelli, in cascate di bootstrapping reciproco.

I modelli markoviani, in particolare quando estesi attraverso framework Bayesiani, HMM gerarchici, e integrazione con altri sistemi di rappresentazione, forniscono strumenti matematici precisi per formalizzare e testare ipotesi su questi processi. La convergenza tra evidenze comportamentali (dai laboratori di psicologia dello sviluppo), neuroscientifiche (da neuroimaging e neurofisiologia), computazionali (da simulazioni e modelli matematici), e linguistiche (da analisi tipologiche cross-linguistiche) sta producendo una comprensione sempre più ricca e integrata di come i bambini compiono la straordinaria impresa di acquisire la loro lingua nativa.

Guardando al futuro, le sfide rimanenti sono formidabili ma stimolanti:

- Come transizioni da sensibilità statistica implicita a conoscenza grammaticale esplicita?
- Come interagiscono precisamente meccanismi dominio-generalisti con specializzazioni linguistiche?
- Come si integrano acquisizione di forma (sintassi, fonologia) con acquisizione di significato (semantica, pragmatica)?
- Come spiegare variazione tipologica massiva tra lingue pur preservando universali di acquisizione?
- Come ricostruire storia evolutiva che ha portato a capacità linguistica umana unica?

Affrontare queste domande richiederà continuata integrazione di metodi sperimentali, computazionali, e teorici. I modelli markoviani, radicati profondamente nella teoria matematica dei processi stocastici ma connessi intimamente a meccanismi cognitivi e

neurale reali, continueranno senza dubbio a giocare ruolo centrale in questa impresa scientifica.

La prospettiva markoviana sull'acquisizione linguistica, fuori che essere riduzionista o impoverente, rivela la bellezza matematica sottostante a uno dei fenomeni più distintamente umani. Essa mostra che principi formali semplici, iterati attraverso sviluppo e amplificati attraverso interazioni sociali e cognitive, possono generare la ricchezza straordinaria e la creatività illimitata del linguaggio umano. In questa sintesi di semplicità matematica e complessità emergente risiede una delle più profonde lezioni che la scienza cognitiva computazionale può offrire alla nostra comprensione della mente umana.

L'acquisizione del linguaggio, vista attraverso modelli markoviani, non è meramente estrazione di statistiche ma piuttosto costruzione attiva di sistemi di rappresentazione strutturati che permettono comunicazione, pensiero, e cultura. È processo che trasforma rumore acustico in significato, variabilità continua in categorie discrete, esperienze frammentate in conoscenze sistematiche. È, in definitiva, processo attraverso cui i bambini diventano pienamente umani, membri competenti di comunità linguistiche e culturali. Comprenderlo scientificamente richiede gli strumenti più sofisticati della matematica, dell'informatica, delle neuroscienze e della linguistica. Apprezzarlo pienamente richiede anche meraviglia di fronte alla straordinaria capacità della mente umana infantile di navigare questo compito formidabile con tale grazia apparente e successo quasi universale.